

B4

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
3 May 2001 (03.05.2001)

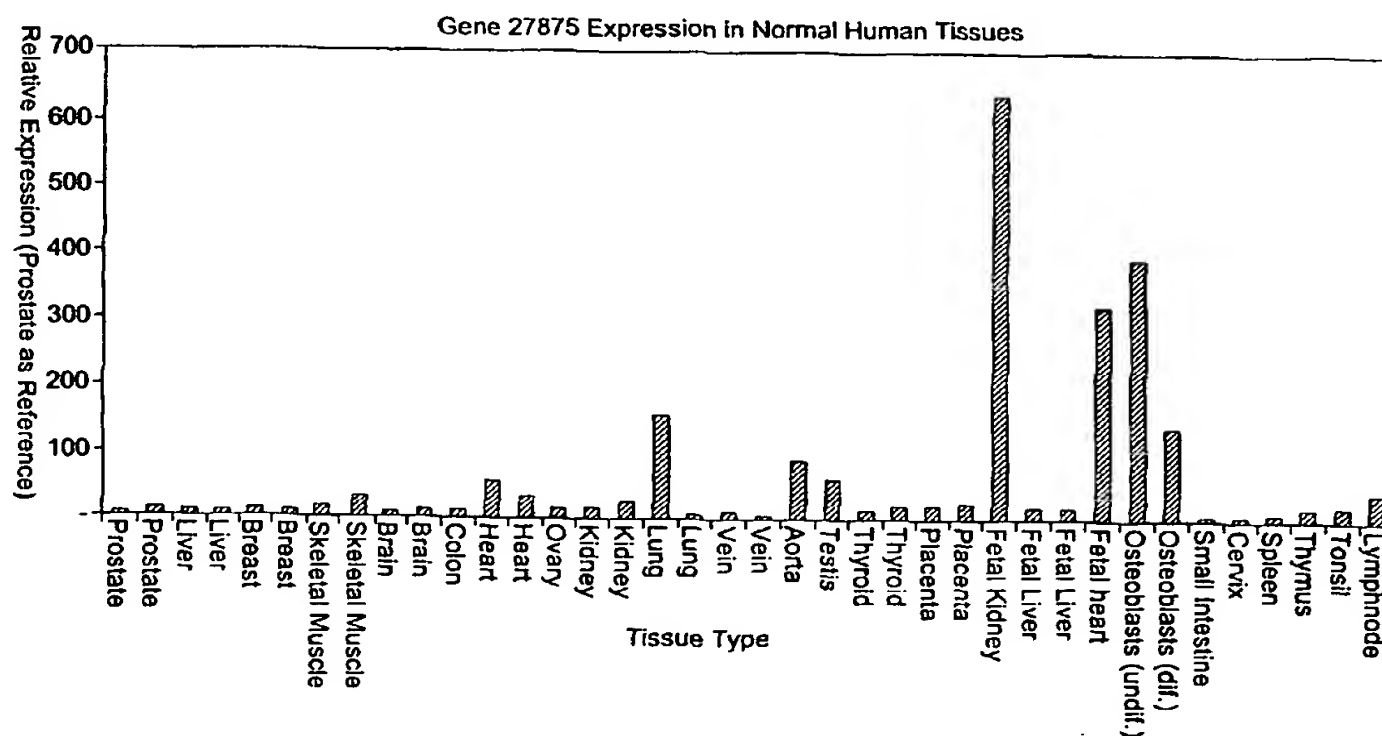
PCT

(10) International Publication Number
WO 01/31034 A1

- (51) International Patent Classification⁷: C12N 15/57, 15/63, 9/64, C07K 16/40, C12Q 1/37
- (21) International Application Number: PCT/US00/29380
- (22) International Filing Date: 25 October 2000 (25.10.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
09/426,282 25 October 1999 (25.10.1999) US
- (71) Applicant (for all designated States except US): MILLENNIUM PHARMACEUTICALS, INC. [US/US]; 75 Sidney Street, Cambridge, MA 02139 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): KAPPELLER-LIBERMANN, Rosana [BR/US]; 86 Beacon Street, Chestnut Hill, MA 02467 (US). WHITE, David [US/US]; 35 Hollingsworth Avenue, Braintree, MA 02184 (US).
- (74) Agents: SPRUILL, W., Murray et al.; Alston & Bird LLP, P.O. Drawer 34009, Charlotte, NC 28234-4009 (US).
- (81) Designated States (national): AE, AG, AL, AM, AT, AT (utility model), AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, CZ (utility model), DE, DE (utility model), DK, DK (utility model), DM, DZ, EE, EE (utility model), ES, FI, FI (utility model), GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (utility model), SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- Published:
— With international search report.

[Continued on next page]

(54) Title: 27875 A HUMAN ADAM-TS HOMOLOG



(57) Abstract: The present invention relates to a newly identified protein, 27875, a human ADAM-TS (A Disintegrin And Metalloproteinase). In particular, the invention relates to 27875 ADAM-TS metalloproteinase polypeptides and polynucleotides, methods of detecting the 27875 ADAM-TS metalloproteinase polypeptides and polynucleotides, and methods of diagnosing and treating 27875 ADAM-TS metalloproteinase-related disorders. Also provided are vectors, host cells, and recombinant methods for making and using the novel molecules.

WO 01/31034 A1

WO 01/31034 A1



For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

27875 A HUMAN ADAM-TS HOMOLOG

FIELD OF THE INVENTION

The present invention relates to a newly identified protein, 27875, a human ADAM-TS (A Disintegrin And Metalloproteinase). In particular, the invention relates to 27875 ADAM-TS metalloproteinase polypeptides and polynucleotides, methods of detecting the 27875 ADAM-TS metalloproteinase polypeptides and polynucleotides, and
5 methods of diagnosing and treating 27875 ADAM-TS metalloproteinase-related disorders. Also provided are vectors, host cells, and recombinant methods for making and using the novel molecules.

BACKGROUND OF THE INVENTION

10

Metalloproteinases are a group of widely distributed proteolytic enzymes that depend on bound Ca^{2+} or Zn^{2+} for activity; however, certain metalloproteinases can readily utilize Mn^{2+} and Mg^{2+} . Biological functions of metalloproteinases include protein maturation, degradation of proteins, such as extracellular matrix proteins,
15 tumor growth, metastasis and angiogenesis.

Disintegrins are integrin ligands that disrupt cell/cell (aggregation) and cell-matrix (adhesion) interactions by inhibiting the binding of other physiological ligands to integrins. Disintegrins have a conserved spacing of cysteine residues that is required for their direct binding to integrin metalloproteinases (Niewiarowski *et al.*
20 (1994) *Semin Hematol* 31:289).

TSP I motifs are conserved domains in Thrombospondin 1 and 2, multifunctional secretory glycoproteins involved in blood clotting, inhibiting angiogenesis and regulating the proliferation, adhesion and migration of normal and tumor cells. The biological activities of thrombospondin 1 and 2 are mediated by the binding of the TSP
25 type I motifs to extracellular matrix molecules, such as heparan sulfate, proteoglycans, fibronectin, laminin and collagen. Thrombospondin-1 is a platelet-derived glycoprotein that is released from platelet alpha granules in response to thrombin

stimulation. It is involved in cell adhesion and modulates cell movement, cell proliferation, neurite outgrowth and angiogenesis.

ADAMs comprise a broad family of multifunctional proteins, characterized as having a disintegrin and metalloproteinase domain (Wolfsberg *et al.* (1995)

5 *Developmental Biol* 169:378-383; Wolfsberg *et al.* (1995) *J Cell Biol* 131:275-278).

Approximately 20 ADAMs have been identified to date. The prototypical ADAM is a membrane-anchored glycoprotein with pro-, metalloproteinase, disintegrin, cystine-rich, epidermal growth factor-like, transmembrane and cytoplasmic domains. Members of the ADAM family of proteins include MDC (ADAM1), fertilin β (ADAM2), cryitestin
10 (ADAM3), epididymal apical protein I, meltrin, MS2, TNF- α converting enzyme, Kuzbanian and metargidin.

ADAMs participate in a variety of roles, including cell-cell and cell-matrix interactions and polypeptide processing. Examples of ADAM functions include tumor cell adhesion (Iba *et al.* (1999) *Am J Pathol* 154:1489-1501), tumor suppression (Emi *et al.* (1993) *Nature Genet* 5:151-157), spermatogenesis and mediation of fusion of gamete
15 membranes (Evans *et al.* (1999) *Biol Reprod* 59:145-152), blastocyst implantation (Olson *et al.* (1998) *Cell Tissue Res* 293:489-498), myotube formation and myoblast fusion (Gilpin *et al.* (1998) *J Biol Chem* 273:157-166), immunity (Higuchi *et al.* (1999) *Immunol Today* 20:278-284), proteolytic processing of ligands that activate epidermal
20 growth factor metalloproteinase (Dong *et al.* (1999) *Proc Natl Acad Sci USA* 96:6235-6240), proteolytic cleavage of Alzheimer's amyloid precursor protein (Lammich *et al.* (1999) *Proc Natl Acad Sci USA* 96:3922-3927; Buxbaum *et al.* (1998) *J Biol Chem* 273:27765-27767), processing of Notch ligands (Qi *et al.* (1999) *Science* 283:91-94), neurogenesis (Rooke *et al.* (1996) *Science* 273:1227-1231), cleavage of murine mannose
25 metalloproteinase to produce a soluble mannose metalloproteinase (Martinez-Pomares *et al.* (1998) *J Biol Chem* 273:23376-23380), and maturation of TNF- α (Lunn *et al.* (1997) *FEBS Lett* 400:333-335). The cell-cell interactions are thought to be mediated by the disintegrin domain.

The cloning of ADAM-TS-1, a novel murine ADAM, was reported (Kuno *et al.* (1997) *J Biol Chem* 272:556-562). ADAM-TS-1 is selectively expressed in the
30 cachexigenic colon 26 adenocarcinoma cell line and is believed to be associated with acute inflammation and cancer cachexia. ADAM-TS-1 is a 951 amino acid polypeptide comprising a signal peptide, a prodomain, a catalytically active zinc-dependent

metalloproteinase domain, a disintegrin domain, and three thrombospondin (TSP) type 1 domains, which are responsible for anchoring ADAM-TS-1 to the extracellular matrix. In contrast to other ADAMs, ADAM-TS-1 does not possess a transmembrane domain or an epidermal growth factor-like domain. Rather, ADAM-TS-1 is secreted and is
5 associated with the extracellular matrix.

More recent reports from this group (Kuno *et al.* (1999) *J. Biol. Chem.* 274:18821-18826; Kuno *et al.* (1998) *J. Biol. Chem.* 273:13912-13917) also showed ADAM-TS-1 to be a unique ADAM family protein with respect to the presence of thrombospondin type 1 motifs and the capacity to bind to the extracellular matrix.
10 Like the other members of the ADAM family, the amino terminal half region of ADAM-TS-1 consists of a proprotein and a metalloproteinase domain and a disintegrin-like domain that share sequence similarity to snake venom metalloproteinases. In contrast, the domain organization of the carboxy terminal half is completely different from other ADAMs. Instead of the transmembrane region,
15 ADAM-TS-1 has three thrombospondin-type 1 motifs found in thrombospondins 1 and 2. These motifs are functional for binding two molecules of heparin. The ADAM-TS-1 is secreted and incorporated into the extracellular matrix. The three thrombospondin-type 1 motifs are responsible for anchoring to the extracellular matrix. The ADAM-TS-1 was shown to have a zinc-binding motif in the
20 metalloproteinase domain providing the capacity to bind to α_2 -macroglobulin. Accordingly, soluble ADAM-TS-1 was shown to be able to form a covalent binding complex with α_2 -macroglobulin. A point mutation in this motif was shown to eliminate the capacity to bind to the α_2 -macroglobulin. In addition, the studies reported that the removal of the prodomain from the ADAM-TS-1 precursor was impaired in a furin-
25 deficient cell line and that the processing ability of the cells was restored by coexpression of the furin cDNA. These results provided evidence that the ADAM-TS-1 precursor is processed *in vivo* by furin endopeptidase in the secretory pathway. It was accordingly proposed that ADAM-TS-1 plays a role in the inflammatory process through its protease activity.

30 Expression of the gene was shown to be induced in kidney and in heart by *in vivo* administration of lipopolysaccharide, suggesting a possible role in the inflammatory reaction. (Kuno *et al.* (1998)).

Using a transient expression system, it was shown that both precursor and

processed forms of ADAM-TS-1 are secreted from cells. The majority was associated with the extracellular matrix. When cells were cultured in the presence of heparin, the mature form of ADAM-TS-1 was detected in cell culture medium, suggesting that the binding of the protein to the extracellular matrix is mediated through a sulfated glycosaminoglycan. Deletion mutation analysis showed that the spacer region and the three thrombospondin-type 1 motifs in the carboxy terminal region are important for interaction with the extracellular matrix (Kuno *et al.* (1998)).

The thrombospondin-type 1 motif is conserved in thrombospondins 1 and 2 which are multifunctional extracellular matrix proteins that influence cell adhesion, motility, and growth (Kuno *et al.* (1998)). Thrombospondin-type 1 motifs and thrombospondins have two conserved heparin-binding segments: W(S/G)XWSXW (SEQ ID NO: 5) and CSVTCG (SEQ ID NO:6). ADAM-TS-1 contains a middle thrombospondin 1 motif with sequences similar to the following heparin-binding segments in thrombospondins: WGPWGPW (SEQ ID NO:7) and CS(R/K)TCG (SEQ ID NO:8). The carboxy terminal submotifs have only the latter sequence. Kuno *et al.* (1998) show that the middle and carboxy terminal TSP submotifs of the ADAM-TS-1 protein are able to bind heparin. The report concluded that the data demonstrate that the interaction between the three motifs and sulfated glycosaminoglycans in the extracellular matrix, such as heparan sulfate, plays a role in the extracellular matrix binding of the ADAM-TS protein. However, the report also showed that truncation of the spacer region intervening between the middle and carboxyl terminal TSP-type 1 motifs significantly reduced the extracellular matrix binding of the protein. Accordingly, it was concluded that, in addition to the three TSP Type 1 motifs, the carboxy terminal spacer domain is important for tight binding to the extracellular matrix. Finally, the report showed that the protein is associated with the extracellular matrix through multiple independent extracellular matrix attachment sites in the carboxy terminal region.

Within the proprotein domain, there are two cleavage sites (RRRR, 178-182, SEQ ID NO:9) (RKRR, 233-236, SEQ ID NO:10) for the furin-like protease. Furin cleaves a wide variety of precursor proteins at the consensus sequence RX(K/R)R (SEQ ID NO:11). Furin cleavage sites are found in a number of precursor proteins that are transported to the cell surface. (Kuno *et al.* (1998)). The ADAM-TS-1 protein has a zinc-binding motif (HEXXH) (SEQ ID NO:12) in its metalloproteinase domain. Accordingly, it was suggested that this protein is secreted from cells as a proteolytically

active form by cleavage with a furin-like enzyme.

Tortorella *et al.* ((1999) *Science* 284:1664-1666) purified the metalloproteinase aggrecanase-1 (ADAM-TS-4) from IL-1-stimulated bovine nasal cartilage conditioned medium and then cloned and expressed the human ortholog. This protease represents a
5 cartilage aggrecanase that cleaves aggrecan at the Glu³⁷³-Ala³⁷⁴ bond to produce fragments similar to those found in the synovial fluid of patients with various types of arthritis. This recombinant molecule provides a target for development of therapeutics to prevent the loss of articular cartilage in arthritis. Aggrecan degradation is an important factor in the erosion of articular cartilage in arthritic diseases. The degradation
10 involves proteolysis in the core protein near the amino terminus where two major cleavage sites have been identified. One of these is the Glu³⁷³-Ala³⁷⁴ cleavage site. Aggrecan fragments cleaved from this site have been identified in cultures undergoing cartilage matrix degradation and in arthritic synovial fluids. Incubation of purified aggrecanase-1 with bovine aggrecan produced fragments generated by cleavage at this
15 site. The fragments were identified by an assay using the neoepitope antibody, BC-3, to detect products formed by specific cleavage at this bond. Further, including SF775, a potent aggrecanase inhibitor, blocked binding of the aggrecanase to a specific inhibitor resin.

The amino terminal and two internal sequences of bovine aggrecanase 1 were
20 found to be 50 to 60% identical to the inflammation-associated murine protein ADAM-TS-1. The aggrecanase 1 contains a signal sequence followed by a propeptide domain with a potential cysteine switch at Cys¹⁹⁴ and a potential furin cleavage site that precedes the catalytic domain. The catalytic domain has a zinc-binding motif similar to the HEXXHXXGXXH (SEQ ID NO:13) motif found in matrix metalloproteinases and
25 ADAMs. The enzyme also contains a disintegrin-like domain and lacks the transmembrane domain and cytoplasmic tail present in many ADAMs. It ends with a carboxy terminal domain that contains a thrombospondin-type 1 motif similar to those present in ADAM-TS-1. It is likely synthesized as a zymogen that is cleaved to remove the propeptide domain to generate the mature active enzyme. A compound that
30 interferes with the normal pro-MMP activation through a cysteine switch mechanism inhibits cleavage of aggrecan in cartilage organ cultures. The enzyme was shown to be ineffective in cleaving several substrates that are cleaved by matrix metalloproteinases including the extracellular matrix molecules type II collagen, thrombospondin, and

15 response. Although not designated as ADAM-TS proteins, the proteins are clearly
members of the ADAM-TS family, containing metalloproteinase, disintegrin, and
thrombospondin domains. In fact, the reference indicates that the mouse homolog of one
of the cloned genes is the ADAM-TS-1. The report also refers to pNP-1 (procollagenase
1 N-proteinase) having a structural resemblance and high sequence similarity to both of
20 the cloned METH proteins. The reference cites Colige *et al.* (*Proc. Natl. Acad. Sci. USA*
94:2374-2379 (1997)) for the identification of this new protein. The authors discussed
the two proteins as novel inhibitors of angiogenesis. They cited four additional members
of the family represented as partial ESTs. The authors also pointed out that despite the
identical structure and the high levels of amino acid similarities in the two proteins, the
25 pattern of expression differs significantly. It was suggested that the differences are most
likely the result of specific cis-acting elements in the non-coding regulatory sequences.
It was proposed that proteins with similar or identical function, but different tissue
specificity, may participate as specific angiogenic inhibitors regulating vascular
networks in different organs or in specific physiological responses. Alternatively, it was
30 proposed that small differences in sequence might confer significant differences in tissue
specificity. Further, whereas ADAM-TS-1 was identified in a screen of genes associated
with the induction of cachexia and appears to be regulated by inflammatory cytokines,
the METH-2 is not reported to have these features. Finally, the authors discussed the

disintegrin motif present in both proteins. The disintegrin motif can contain an RGD (or RGX) motif with a negatively charged residue at the X-position. This sequence binds two integrins and serves as ligand or an antagonist of ligand binding. The authors pointed out that inactivation of integrins with antibodies has been shown to inhibit
5 neovascularization during development and in tumorigenesis.

Abbaszade *et al.* ((1999) *J. Biol. Chem.* 274:23443-23450)) report the cloning and characterization of a second aggrecanase, designated ADAM-TS-11. It was shown to have extensive homology to ADAM-TS-4 (aggrecanase-1) and to ADAM-TS-1. The recombinant human ADAM-TS-11 was expressed in insect cells and shown to cleave
10 aggrecan at the Glu-Ala site. Aggrecan is the major proteoglycan of cartilage and is responsible for its compressibility and stiffness. Results from several studies cited by the authors suggest that the cleavage at the Glu-Ala site is responsible for increased aggrecan degradation observed in inflammatory joint disease. Gene expression of both the ADAM-TS-4 and ADAM-TS-1 were examined in a variety of normal and arthritic
15 human tissues. ADAM-TS-1 was shown to be highly expressed in arthritic fibrous tissues and arthritic joint capsule. The ADAM-TS-4 and ADAM-TS-11 both showed moderate expression in arthritic fibrous tissue and arthritic joint capsule. However, expression was not limited to these tissues alone. The ADAM-TS-11 appears to be synthesized in an inactive pro form. The N-terminal peptide sequence of the enzyme
20 purified from bovine-cartilage-conditioned medium starts immediately C terminal of the consensus furin cleavage site. Accordingly, the inhibition of furin can block aggrecan cleavage.

Accordingly, ADAMs and ADAM-TSs are a major target for drug action and development. Therefore, it is valuable to the field of pharmaceutical development to
25 identify and characterize previously unknown ADAMs and ADAM-TSs. The present invention advances the state of the art by providing a previously unidentified human ADAM-TS having 39% sequence identity and 67% sequence similarity with murine ADAM-TS-1 and a second human metalloproteinase with homology to the ADAM-TS family, and especially high homology to the above novel ADAM-TS.

SUMMARY OF THE INVENTION

A novel ADAM-TS cDNA, 27875 metalloproteinase, and the deduced 27875
5 metalloproteinase polypeptide are described herein. Accordingly, the invention
provides isolated 27875 metalloproteinase nucleic acid molecules having the sequence
shown in SEQ ID NO:2, and variants and fragments thereof.

It is also an object of the invention to provide nucleic acid molecules encoding
the 27875 metalloproteinase polypeptide, and variants and fragments thereof. Such
10 nucleic acid molecules are useful as targets and reagents in 27875 metalloproteinase
expression assays, are applicable to treatment and diagnosis of 27875 metalloproteinase-
related disorders and are useful for producing novel 27875 metalloproteinase
polypeptides by recombinant methods.

The invention also provides a partial cDNA and deduced amino acid sequence
15 for a second human metalloproteinase with homology to the ADAM-TS family, and
particularly high homology to the 27875 metalloproteinase. This protein has been
designated 42812. Further, where appropriate, although the disclosure herein and all
embodiments are explicitly directed to the 27875 metalloproteinase, these embodiments
apply as well to the 42812 metalloproteinase protein. An alignment between these two
20 proteins is shown herein.

The invention thus further provides nucleic acid constructs comprising the
nucleic acid molecules described herein. In a preferred embodiment, the nucleic acid
molecules of the invention are operatively linked to a regulatory sequence. The
invention also provides vectors and host cells for expressing the 27875 metalloproteinase
25 nucleic acid molecules and polypeptides, and particularly recombinant vectors and host
cells.

In another aspect, it is an object of the invention to provide isolated 27875
metalloproteinase polypeptides and fragments and variants thereof, including a
polypeptide having the amino acid sequence shown in SEQ ID NO:1. The disclosed
30 27875 metalloproteinase polypeptides are useful as reagents or targets in 27875
metalloproteinase assays and are applicable to treatment and diagnosis of 27875
metalloproteinase-related disorders.

The invention also provides assays for determining the activity of or the presence or absence of the 27875 metalloproteinase polypeptides or nucleic acid molecules in a biological sample, including for disease diagnosis. In addition, the invention provides assays for determining the presence of a mutation in the polypeptides or nucleic acid
5 molecules, including for disease diagnosis.

A further object of the invention is to provide compounds that modulate expression of the 27875 metalloproteinase for treatment and diagnosis of 27875 metalloproteinase-related disorders. Such compounds may be used to treat conditions related to aberrant activity or expression of the 27875 metalloproteinase polypeptides or
10 nucleic acids.

The disclosed invention further relates to methods and compositions for the study, modulation, diagnosis and treatment of 27875 metalloproteinase related disorders. The compositions include 27875 metalloproteinase polypeptides, nucleic acids, vectors, transformed cells and related variants thereof. In particular, the
15 invention relates to the diagnosis and treatment of 27875 metalloproteinase-related disorders of bone, lung, heart, skeletal muscle, aorta, testis, and kidney, and more specifically of bone. Since the gene is highly expressed in undifferentiated osteoblasts, the invention even more specifically relates to disorders involving osteoblast function, growth, and differentiation, and to modulation of gene expression
20 in osteoblasts. Accordingly, specific disorders include, but are not limited to, osteoporosis and osteopetrosis.

In yet another aspect, the invention provides antibodies or antigen-binding fragments thereof that selectively bind the 27875 metalloproteinase polypeptides and fragments. Such antibodies and antigen binding fragments have use in the detection of
25 the 27875 metalloproteinase polypeptide, and in the prevention, diagnosis and treatment of 27875 metalloproteinase related disorders.

DESCRIPTION OF THE DRAWINGS

30 Figure 1 shows the 27875 metalloproteinase cDNA sequence (SEQ ID NO:2) and the deduced amino acid sequence (SEQ ID NO:1).

Figure 2 shows a 27875 metalloproteinase hydrophobicity plot.

N-glycosylation sites, the actual modified residue is the first amino acid. For cAMP- and cGMP-dependent protein kinase phosphorylation sites, the actual modified
10 residue is the first amino acid. For casein kinase II phosphorylation sites, the actual modified residue is the first amino acid. For tyrosine kinase phosphorylation sites, the actual modified residue is the last amino acid. For N-myristoylation sites, the actual modified residue is the first amino acid

15 Figure 5 shows relative expression of 27875 metalloproteinase mRNA in normal human tissues.

Figure 6 shows *in situ* hybridization of an 27875 metalloproteinase probe with human fetal bone.

20

Figure 5 shows a schematic of the various modules in some ADAM-TS family members, including the 27875 metalloproteinase.

Figure 8 shows the nucleotide (SEQ ID NO:4) and deduced amino acid (SEQ ID
25 NO:3) sequence for the partial cDNA designated 42812 ADAM-TS.

Figure 9 shows an alignment of the 42812 (SEQ ID NO:3) and 27875 (SEQ ID NO: 1) amino acid sequences.

30

DETAILED DESCRIPTION OF THE INVENTION

The growth, development and maintenance of bone is a highly regulated process. Bone mass reflects the balance of bone formation and resorption which at the cellular

level involves the coordinate regulation of bone forming (osteoblast) and bone resorbing (osteoclast) cells. Each of these cell types is influenced by a wide variety of hormones, inflammatory mediators and growth factors. Importantly, osteoblast-derived secreted factors are known regulators of osteoclast formation and/or activity *in vivo*.

- 5 Accordingly, it would be beneficial to identify these osteoblast-secreted factors. Such factors may function to regulate osteoblast activity including both cytokine and hormone processing as well as extracellular matrix homeostasis. Modulation of the activity of such factors (for example, via the use of small molecule inhibitors) may prove beneficial for blocking activities of osteoblasts that are associated with accelerated osteoclast
10 formation/activities and subsequent bone resorptive function.

The invention is based on the identification of the novel human ADAM-TS 27875 metalloproteinase, which is expressed at high levels in undifferentiated osteoblast, fetal heart and fetal kidney. The 27875 metalloproteinase cDNA was identified based on consensus motifs or protein domains characteristic of the ADAM-
15 TS family of metalloproteases. Specifically, a novel human gene, termed the 27875 metalloproteinase, is provided. This sequence, and other nucleotide sequences encoding the 27875 metalloproteinase protein or fragments and variants thereof, are referred to as "27875 metalloproteinase sequences".

The 27875 metalloproteinase cDNA was identified in a human bone cell cDNA
20 library. Specifically, an expressed sequence tag (EST) found in a human bone library was selected based on homology to known ADAM-TS sequences. Based on this EST sequence, primers were designed to identify a full length clone from a human bone cDNA library. Positive clones were sequenced and the overlapping fragments were assembled. The 27875 metalloproteinase amino acid sequence is shown in Figure 1 and
25 SEQ ID NO:1. The 27875 metalloproteinase cDNA sequence is shown in Figure 1 and SEQ ID NO:2.

Analysis of the assembled sequence revealed that the cloned cDNA molecule encodes an ADAM-TS-like polypeptide. BLAST analysis indicated that the 27875 metalloproteinase protein displays closest similarity to the murine ADAM-TS-1
30 protein, with approximately 39% identity and 67% overall similarity, indicating that the 27875 metalloproteinase is the human ortholog of this murine protein.

The 27875 metalloproteinase sequence of the invention belongs to the ADAM-TS family of molecules having conserved functional features. The term

NO:2), which encodes an 1687 amino acid protein (SEQ ID NO:1). The 27875 metalloproteinase contains a metalloproteinase domain at residues 244-259 of SEQ ID NO:1 and a disintegrin domain at residues 541-592 of SEQ ID NO:1. A zinc-binding domain (active site) is found at approximately amino acids 385-394. The protein also
15 contains 5 thrombospondin domains located from about amino acid 542-592, 825-868, 949-988, 1415-1463, and 1466-1521. SignalP (eukaryote) analysis of the amino terminal 70 amino acids of the polypeptide predicts a 30 amino acid signal peptide, which is cleaved to produce the mature 27875 metalloproteinase polypeptide (residues 31-1687 of SEQ ID NO:1).

20 Prosites program analysis was used to predict various sites within the 27875 metalloproteinase protein. N-glycosylation sites were predicted at about amino acid residues 94-97, 693-696, 778-781, 950-953, 971-974, 1412-1415, 1419-1422 and 1470 to 1473 of SEQ ID NO:1. A glycosaminoglycan attachment site was predicted at about amino acid residues 1006-1009 of SEQ ID NO:1. cAMP- and cGMP-dependent protein
25 kinase phosphorylation sites were predicted at amino acid residues 872-875 and 1606-1609 of SEQ ID NO:1. Protein kinase C phosphorylation sites were predicted at amino acid residues 6-8, 73-75, 110-112, 214-216, 313-315, 342-344, 569-571, 598-600, 901-903, 962-964, 1035-1037, 1370-1372, 1385-1387, 1440-1442, 1483-1485, 1528-1530, 1599-1601, 1620-1622, 1649-1651 and 1660-1662 of SEQ ID NO:1. Casein kinase II
30 phosphorylation sites were predicted at amino acid residues 147-150, 159-162, 214-217, 342-345, 373-376, 401-404, 505-508, 605-608, 703-706, 917-920, 957-960, 1011-1014, 1192-1195, 1308-1311, 1397-1400, 1440-1443, 1483-1486, 1528-1531 and 1546-1549 of SEQ ID NO:1. A tyrosine kinase phosphorylation site was predicted at amino acid

residues 740-747 of SEQ ID NO:1. N-myristoylation sites were predicted at amino acid residues 55-60, 115-120, 141-146, 379-384, 479-484, 513-518, 539-544, 557-562, 614-619, 667-672, 688-693, 716-721, 765-770, 774-779, 1005-1010, 1039-1044, 1263-1252, 1263-1268, 1358-1363, 1517-1522, 1592-1597 and 1625-1630 of SEQ ID NO:1. An
5 amidation site was predicted at amino acid residues 408-411 of SEQ ID NO:1. A cell attachment sequence was predicted at amino acid residues 195-197 of SEQ ID NO:1. A zinc binding domain is predicted at residues 385 to 394 of SEQ ID NO:1. A Cytochrome C family heme-binding site was predicted at amino acid residues 687-692 of SEQ ID NO:1. A crystallins beta and gamma Greek key motif is predicted at amino
10 acid residues 78-93 of SEQ ID NO:1. A growth factor and cytokine metalloproteinase family signature 2 domain was predicted at amino acid residues 539-545 of SEQ ID NO:1. Thrombospondin domains were predicted by HMMer, Version 2, at amino acid residues 488-567, 542-592, 825-879, 949-994, 1415-1463 and 1466-1521 of SEQ ID NO:1.

15 Northern blot analysis of 27875 metalloproteinase expression in human tissues shows high level expression in cells of osteoblast lineage (Figure 5). A transcript of approximately 4 kb was detected in osteoblast-derived polyA⁺ RNA (not shown). *In situ* hybridization with human fetal bone also showed significant levels of expression in mature and stromal osteoblast progenitors (Figure 6). High 27875
20 metalloproteinase expression was also detected in human fetal kidney and fetal heart (Figure 5). The gene is also significantly expressed in human adult skeletal muscle, heart, lung, aorta, testes, and lymph node (Figure 5) as well as in thymus and normal foreskin melanocytes (not shown).

Expression of 27875 metalloproteinase mRNA in the above cells and tissues
25 indicates that the 27875 metalloproteinase is likely to be involved in the proper function and in disorders of these tissues, especially the bone, where the gene is expressed in osteoblasts. Accordingly, the disclosed invention further relates to methods and compositions for the study, modulation, diagnosis and treatment of 27875 metalloproteinase related disorders, especially disorders of the bone that
30 include, but are not limited to, osteoporosis and osteopetrosis. Since the gene is expressed in undifferentiated osteoblasts, disorders related to osteoblast production, function, and differentiation are particularly relevant to the invention. The compositions include 27875 metalloproteinase polypeptides, nucleic acids, vectors,

transformed cells and related variants and fragments thereof, as well as agents that modulate expression of the polypeptides and polynucleotides. In particular, the invention relates to the modulation, diagnosis and treatment of 27875 metalloproteinase related disorders as described herein.

- 5 Disorders involving the lung include, but are not limited to, congenital anomalies; atelectasis; diseases of vascular origin, such as pulmonary congestion and edema, including hemodynamic pulmonary edema and edema caused by microvascular injury, adult respiratory distress syndrome (diffuse alveolar damage), pulmonary embolism, hemorrhage, and infarction, and pulmonary hypertension and
- 10 vascular sclerosis; chronic obstructive pulmonary disease, such as emphysema, chronic bronchitis, bronchial asthma, and bronchiectasis; diffuse interstitial (infiltrative, restrictive) diseases, such as pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia (pulmonary infiltration with eosinophilia),
- 15 *Bronchiolitis obliterans*-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, including Goodpasture syndrome, idiopathic pulmonary hemosiderosis and other hemorrhagic syndromes, pulmonary involvement in collagen vascular disorders, and pulmonary alveolar proteinosis; complications of therapies, such as drug-induced lung disease, radiation-induced lung disease, and lung transplantation;
- 20 tumors, such as bronchogenic carcinoma, including paraneoplastic syndromes, bronchioloalveolar carcinoma, neuroendocrine tumors, such as bronchial carcinoid, miscellaneous tumors, and metastatic tumors; pathologies of the pleura, including inflammatory pleural effusions, noninflammatory pleural effusions, pneumothorax, and pleural tumors, including solitary fibrous tumors (pleural fibroma) and malignant
- 25 mesothelioma.

- Disorders involving the heart, include but are not limited to, heart failure, including but not limited to, cardiac hypertrophy, left-sided heart failure, and right-sided heart failure; ischemic heart disease, including but not limited to angina pectoris, myocardial infarction, chronic ischemic heart disease, and sudden cardiac death;
- 30 hypertensive heart disease, including but not limited to, systemic (left-sided) hypertensive heart disease and pulmonary (right-sided) hypertensive heart disease; valvular heart disease, including but not limited to, valvular degeneration caused by calcification, such as calcific aortic stenosis, calcification of a congenitally bicuspid

aortic valve, and mitral annular calcification, and myxomatous degeneration of the mitral valve (mitral valve prolapse), rheumatic fever and rheumatic heart disease, infective endocarditis, and noninfected vegetations, such as nonbacterial thrombotic endocarditis and endocarditis of systemic lupus erythematosus (Libman-Sacks disease), carcinoid heart disease, and complications of artificial valves; myocardial disease, including but not limited to dilated cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy, and myocarditis; pericardial disease, including but not limited to, pericardial effusion and hemopericardium and pericarditis, including acute pericarditis and healed pericarditis, and rheumatoid heart disease; neoplastic heart disease, including but not limited to, primary cardiac tumors, such as myxoma, lipoma, papillary fibroelastoma, rhabdomyoma, and sarcoma, and cardiac effects of noncardiac neoplasms; congenital heart disease, including but not limited to, left-to-right shunts--late cyanosis, such as atrial septal defect, ventricular septal defect, patent ductus arteriosus, and atrioventricular septal defect, right-to-left shunts--early cyanosis, such as tetralogy of fallot, transposition of great arteries, truncus arteriosus, tricuspid atresia, and total anomalous pulmonary venous connection, obstructive congenital anomalies, such as coarctation of aorta, pulmonary stenosis and atresia, and aortic stenosis and atresia, and disorders involving cardiac transplantation.

Disorders involving the skeletal muscle include tumors such as rhabdomyosarcoma.

Disorders involving blood vessels include, but are not limited to, responses of vascular cell walls to injury, such as endothelial dysfunction and endothelial activation and intimal thickening; vascular diseases including, but not limited to, congenital anomalies, such as arteriovenous fistula, atherosclerosis, and hypertensive vascular disease, such as hypertension; inflammatory disease--the vasculitides, such as giant cell (temporal) arteritis, Takayasu arteritis, polyarteritis nodosa (classic), Kawasaki syndrome (mucocutaneous lymph node syndrome), microscopic polyanglitis (microscopic polyarteritis, hypersensitivity or leukocytoclastic angitis), Wegener granulomatosis, thromboanglitis obliterans (Buerger disease), vasculitis associated with other disorders, and infectious arteritis; Raynaud disease; aneurysms and dissection, such as abdominal aortic aneurysms, syphilitic (luetic) aneurysms, and aortic dissection (dissecting hematoma); disorders of veins and lymphatics, such as varicose veins, thrombophlebitis and phlebothrombosis, obstruction of superior vena

cava (superior vena cava syndrome), obstruction of inferior vena cava (inferior vena cava syndrome), and lymphangitis and lymphedema; tumors, including benign tumors and tumor-like conditions, such as hemangioma, lymphangioma, glomus tumor (glomangioma), vascular ectasias, and bacillary angiomatosis, and intermediate-grade (borderline low-grade malignant) tumors, such as Kaposi sarcoma and hemangioendothelioma, and malignant tumors, such as angiosarcoma and hemangiopericytoma; and pathology of therapeutic interventions in vascular disease, such as balloon angioplasty and related techniques and vascular replacement, such as coronary artery bypass graft surgery.

Disorders involving the testis and epididymis include, but are not limited to, congenital anomalies such as cryptorchidism, regressive changes such as atrophy, inflammations such as nonspecific epididymitis and orchitis, granulomatous (autoimmune) orchitis, and specific inflammations including, but not limited to, gonorrhea, mumps, tuberculosis, and syphilis, vascular disturbances including torsion, testicular tumors including germ cell tumors that include, but are not limited to, seminoma, spermatocytic seminoma, embryonal carcinoma, yolk sac tumor choriocarcinoma, teratoma, and mixed tumors, tumor of sex cord-gonadal stroma including, but not limited to, leydig (interstitial) cell tumors and sertoli cell tumors (androblastoma), and testicular lymphoma, and miscellaneous lesions of tunica vaginalis.

Disorders involving the kidney include, but are not limited to, congenital anomalies including, but not limited to, cystic diseases of the kidney, that include but are not limited to, cystic renal dysplasia, autosomal dominant (adult) polycystic kidney disease, autosomal recessive (childhood) polycystic kidney disease, and cystic diseases of renal medulla, which include, but are not limited to, medullary sponge kidney, and nephronophthisis-uremic medullary cystic disease complex, acquired (dialysis-associated) cystic disease, such as simple cysts; glomerular diseases including pathologies of glomerular injury that include, but are not limited to, in situ immune complex deposition, that includes, but is not limited to, anti-GBM nephritis, Heymann nephritis, and antibodies against planted antigens, circulating immune complex nephritis, antibodies to glomerular cells, cell-mediated immunity in glomerulonephritis, activation of alternative complement pathway, epithelial cell injury, and pathologies involving mediators of glomerular injury including cellular and soluble mediators, acute

glomerulonephritis, such as acute proliferative (poststreptococcal, postinfectious) glomerulonephritis, including but not limited to, poststreptococcal glomerulonephritis and nonstreptococcal acute glomerulonephritis, rapidly progressive (crescentic) glomerulonephritis, nephrotic syndrome, membranous glomerulonephritis (membranous nephropathy), minimal change disease (lipoid nephrosis), focal segmental glomerulosclerosis, membranoproliferative glomerulonephritis, IgA nephropathy (Berger disease), focal proliferative and necrotizing glomerulonephritis (focal glomerulonephritis), hereditary nephritis, including but not limited to, Alport syndrome and thin membrane disease (benign familial hematuria), chronic glomerulonephritis, glomerular lesions associated with systemic disease, including but not limited to, systemic lupus erythematosus, Henoch-Schönlein purpura, bacterial endocarditis, diabetic glomerulosclerosis, amyloidosis, fibrillary and immunotactoid glomerulonephritis, and other systemic disorders; diseases affecting tubules and interstitium, including acute tubular necrosis and tubulointerstitial nephritis, including but not limited to, pyelonephritis and urinary tract infection, acute pyelonephritis, chronic pyelonephritis and reflux nephropathy, and tubulointerstitial nephritis induced by drugs and toxins, including but not limited to, acute drug-induced interstitial nephritis, analgesic abuse nephropathy, nephropathy associated with nonsteroidal anti-inflammatory drugs, and other tubulointerstitial diseases including, but not limited to, urate nephropathy, hypercalcemia and nephrocalcinosis, and multiple myeloma; diseases of blood vessels including benign nephrosclerosis, malignant hypertension and accelerated nephrosclerosis, renal artery stenosis, and thrombotic microangiopathies including, but not limited to, classic (childhood) hemolytic-uremic syndrome, adult hemolytic-uremic syndrome/thrombotic thrombocytopenic purpura, idiopathic HUS/TTP, and other vascular disorders including, but not limited to, atherosclerotic ischemic renal disease, atheroembolic renal disease, sickle cell disease nephropathy, diffuse cortical necrosis, and renal infarcts; urinary tract obstruction (obstructive uropathy); urolithiasis (renal calculi, stones); and tumors of the kidney including, but not limited to, benign tumors, such as renal papillary adenoma, renal fibroma or hamartoma (renomedullary interstitial cell tumor), angiomyolipoma, and oncocytoma, and malignant tumors, including renal cell carcinoma (hypernephroma, adenocarcinoma of kidney), which includes urothelial carcinomas of renal pelvis.

Bone-forming cells include the osteoprogenitor cells, osteoblasts, and

osteocytes. The disorders of the bone are complex because they may have an impact on the skeleton during any of its stages of development. Hence, the disorders may have variable manifestations and may involve one, multiple or all bones of the body. Such disorders include, congenital malformations, achondroplasia and thanatophoric dwarfism, diseases associated with abnormal matrix such as type 1 collagen disease, osteoporosis, Paget disease, rickets, osteomalacia, high-turnover osteodystrophy, low-turnover aplastic disease, osteonecrosis, pyogenic osteomyelitis, tuberculous osteomyelitis, osteoma, osteoid osteoma, osteoblastoma, osteosarcoma, osteochondroma, chondromas, chondroblastoma, chondromyxoid fibroma, chondrosarcoma, fibrous cortical defects, fibrous dysplasia, fibrosarcoma, malignant fibrous histiocytoma, Ewing sarcoma, primitive neuroectodermal tumor, giant cell tumor, and metastatic tumors.

Disorders involving the thymus include developmental disorders, such as DiGeorge syndrome with thymic hypoplasia or aplasia; thymic cysts; thymic hypoplasia, which involves the appearance of lymphoid follicles within the thymus, creating thymic follicular hyperplasia; and thymomas, including germ cell tumors, lymphomas, Hodgkin disease, and carcinoids. Thymomas can include benign or encapsulated thymoma, and malignant thymoma Type I (invasive thymoma) or Type II, designated thymic carcinoma.

The sequences of the invention find use in diagnosis of disorders involving an increase or decrease in 27875 metalloproteinase expression relative to normal expression, such as a proliferative disorder, a differentiative disorder, or a developmental disorder. The sequences also find use in modulating 27875 metalloproteinase-related responses. By "modulating" is intended the upregulating or downregulating of a response. That is, the compositions of the invention affect the targeted activity in either a positive or negative fashion.

27875 Metalloproteinase Polypeptides

The invention relates to the novel 27875 metalloproteinase, having the deduced amino acid sequence shown in Figure 1 (SEQ ID NO:1).

Thus, present invention provides an isolated or purified 27875 metalloproteinase polypeptide and variants and fragments thereof. "27875 metalloproteinase polypeptide" or "27875 metalloproteinase protein" refers to the polypeptide in SEQ ID NO:1. The

term "27875 metalloproteinase protein" or "27875 metalloproteinase polypeptide", however, further includes the numerous variants described herein, as well as fragments derived from the full-length 27875 metalloproteinase and variants.

27875 metalloproteinase polypeptides can be purified to homogeneity. It is understood, however, that preparations in which the polypeptide is not purified to homogeneity are useful and considered to contain an isolated form of the polypeptide. The critical feature is that the preparation allows for the desired function of the polypeptide, even in the presence of considerable amounts of other components. Thus, the invention encompasses various degrees of purity.

As used herein, a polypeptide is said to be "isolated" or "purified" when it is substantially free of cellular material when it is isolated from recombinant and non-recombinant cells, or free of chemical precursors or other chemicals when it is chemically synthesized. A polypeptide, however, can be joined to another polypeptide with which it is not normally associated in a cell and still be considered "isolated" or "purified."

In one embodiment, the language "substantially free of cellular material" includes preparations of 27875 metalloproteinase having less than about 30% (by dry weight) other proteins (i.e., contaminating protein), less than about 20% other proteins, less than about 10% other proteins, or less than about 5% other proteins. When the polypeptide is recombinantly produced, it can also be substantially free of culture medium, i.e., culture medium represents less than about 20%, less than about 10%, or less than about 5% of the volume of the protein preparation.

The 27875 metalloproteinase polypeptide is also considered to be isolated when it is part of a membrane preparation or is purified and then reconstituted with membrane vesicles or liposomes.

The language "substantially free of chemical precursors or other chemicals" includes preparations of the 27875 metalloproteinase polypeptide in which it is separated from chemical precursors or other chemicals that are involved in its synthesis. The language "substantially free of chemical precursors or other chemicals" includes, but is not limited to, preparations of the polypeptide having less than about 30% (by dry weight) chemical precursors or other chemicals, less than about 20% chemical precursors or other chemicals, less than about 10% chemical precursors or other chemicals, or less than about 5% chemical precursors or other chemicals.

In one embodiment, the 27875 metalloproteinase polypeptide comprises the amino acid sequence shown in SEQ ID NO:1. However, the invention also encompasses sequence variants. Variants include a substantially homologous protein encoded by the same genetic locus in an organism, i.e., an allelic variant. Variants also encompass proteins derived from other genetic loci in an organism, but having substantial homology to 27875 metalloproteinase of SEQ ID NO:1. Variants also include proteins substantially homologous to 27875 metalloproteinase but derived from another organism, i.e., an ortholog. Variants also include proteins that are substantially homologous to 27875 metalloproteinase that are produced by chemical synthesis.

10 Variants also include proteins that are substantially homologous to 27875 metalloproteinase that are produced by recombinant methods. It is understood, however, that variants exclude any amino acid sequences disclosed prior to the invention.

Preferred 27875 metalloproteinase polypeptides of the present invention have an amino acid sequence sufficiently identical to the amino acid sequence of SEQ ID NO:1. The term "sufficiently identical" is used herein to refer to a first amino acid or nucleotide sequence that contains a sufficient or minimum number of identical or equivalent (e.g., with a similar side chain) amino acid residues or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences have a common structural domain and/or common functional activity. For example, amino acid or nucleotide sequences that contain a common structural domain having at least about 45%, 55%, or 65% identity, preferably 75% identity, more preferably 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity are defined herein as sufficiently identical.

The determination of percent identity between two sequences using the algorithms of Karlin and Altschul (1990) *Proc. Natl. Acad. Sci. USA* 87:2264, modified as in Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5877. Such an algorithm is incorporated into the NBLAST and XBLAST programs of Altschul *et al.* (1990) *J. Mol. Biol.* 215:403. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12, to obtain nucleotide sequences homologous to 27875 metalloproteinase nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3, to obtain amino acid sequences homologous to 27875 metalloproteinase protein molecules of the invention. When utilizing BLAST

programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See <http://www.ncbi.nlm.nih.gov>.

As used herein, two proteins (or a region of the proteins) are substantially homologous when the amino acid sequences are at least about 60-65%, 65-70%, 70-
5 75%, typically at least about 80-85%, and most typically at least about 90-95% or more homologous. A substantially homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid sequence hybridizing to the nucleic acid sequence, or portion thereof, of the sequence shown in SEQ ID NO:2 under stringent conditions as more fully described below.

10 To determine the percent identity of two amino acid sequences or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a
15 reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, or 90% of the length of the reference sequence (e.g., when aligning a second sequence to the amino acid sequences herein having 502 amino acid residues, at least 165, preferably at least 200, more preferably at least 250,
20 even more preferably at least 300, and even more preferably at least 350, 400, 450, and 500 amino acid residues are aligned). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules
25 are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

30 The invention also encompasses polypeptides having a lower degree of identity but having sufficient similarity so as to perform one or more of the same functions performed by 27875 metalloproteinase. Similarity is determined by conservative amino acid substitution, as shown in Table 1. Such substitutions are those that substitute a

given amino acid in a polypeptide by another amino acid of like characteristics.

Conservative substitutions are likely to be phenotypically silent. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie *et al.*, *Science* 247:1306-1310 (1990).

10 TABLE 1. Conservative Amino Acid Substitutions.

Aromatic	Phenylalanine
	Tryptophan
	Tyrosine
Hydrophobic	Leucine
	Isoleucine
	Valine
Polar	Glutamine
	Asparagine
Basic	Arginine
	Lysine
	Histidine
Acidic	Aspartic Acid
	Glutamic Acid
Small	Alanine
	Serine
	Threonine
	Methionine
	Glycine

A variant polypeptide can differ in amino acid sequence by one or more substitutions, deletions, insertions, inversions, fusions, and truncations or a combination of any of these. Variant polypeptides can be fully functional or can lack function in one or more activities. Thus, in the present case, variations can affect the
5 function, for example, of one or more of regions including any of the five thrombospondin domains, the disintegrin domain, zinc-binding domain, metalloproteinase domain, the region containing the propeptide, regulatory regions, other substrate binding regions, regions involved in membrane association, regions involved in post-translational modification, for example, by phosphorylation, and
10 regions that are important for effector function (i.e., agents that act upon the protein, such as pro-peptide cleavage).

Fully functional variants typically contain only conservative variation or variation in non-critical residues or in non-critical regions. Functional variants can also contain substitution of similar amino acids, which results in no change or an insignificant
15 change in function. Alternatively, such substitutions may positively or negatively affect function to some degree.

Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncation or a substitution, insertion, inversion, or deletion in a critical residue or critical region.

20 As indicated, variants can be naturally-occurring or can be made by recombinant means or chemical synthesis to provide useful and novel characteristics for 27875 metalloproteinase polypeptide. This includes preventing immunogenicity from pharmaceutical formulations by preventing protein aggregation.

Useful variations further include alteration of functional activity. For example,
25 one embodiment involves a variation at the substrate peptide binding site that results in binding but not hydrolysis or slower hydrolysis of the peptide substrate. A further useful variation at the same site can result in altered affinity for the peptide substrate. Useful variations also include changes that provide for affinity for another peptide substrate. Useful variations further include the ability to bind integrin with greater or lesser
30 affinity, such as not to bind integrin or to bind integrin but not release it. Further useful variations include alteration in the ability of the propeptide to be cleaved by a cleavage protein, for example, by furin, including alteration in the binding or recognition site. Further, the cleavage site can also be modified so that recognition

and cleavage are by a different protease. A useful variation includes binding, but not cleavage, by such a protease. Further useful variations involve variations in the TSP domain, such as in the ability to bind heparin or other sulfated glycosaminoglycan, such as greater or lesser affinity, or a change in specificity. A further useful variation
5 involves a variation in the ability to be bound by zinc, including a greater or lesser affinity for the metal. Further variation could include a variation in the specificity of metal binding, in other words, the ability to be bound by a different metal ion.

Another useful variation provides a fusion protein in which one or more domains or subregions are operationally fused to one or more domains, subregions, or motifs
10 from another ADAMs-TS or ADAM. For example, the transmembrane domain from an ADAM protein can be introduced into the 27875 ADAM-TS such that the protein is anchored in the cell surface. Other permutations include the number of thrombospondin domains, mixing of thrombospondin domains from different ADAM-TS families, spacer regions (between thrombospondin domains), from different
15 ADAM-TS families, the metalloproteinase domain, the propeptide domain, and the disintegrin domain. Mixing these various domains can allow the formation of novel ADAM-TS molecules with different host cell, substrate, and effector molecule (one that acts on the ADAM-TS) specificity.

The term "substrate" is intended to refer not only to the peptide substrate that
20 is cleaved by the metalloproteinase domain, but to refer to any component with which the 27875 polypeptide interacts in order to produce an effect on that component or a subsequent biological effect that is a result of interacting with that component. This includes, but is not limited to, for example, interaction with extracellular matrix components and integrin. However, it is understood that a substrate also includes
25 peptides that are cleaved as a result of catalysis in the metalloproteinase domain.

Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham *et al.* (1985) *Science* 244:1081-1085). The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant
30 molecules are then tested for biological activity, such as peptide bond hydrolysis *in vitro* or related biological activity, such as proliferative activity. Sites that are critical for binding can also be determined by structural analysis such as crystallization, nuclear

magnetic resonance or photoaffinity labeling (Smith *et al.* (1992) *J. Mol. Biol.* 224:899-904; de Vos *et al.* (1992) *Science* 255:306-312).

The invention thus also includes polypeptide fragments of 27875 metalloproteinase. Fragments can be derived from the amino acid sequence shown in SEQ ID NO. 1. However, the invention also encompasses fragments of the variants of the 27875 metalloproteinase polypeptide as described herein. The fragments to which the invention pertains, however, are not to be construed as encompassing fragments that may be disclosed prior to the present invention.

The longest contiguous stretch of amino acid homology between the 27875 metalloproteinase and ADAM-TS-1 is 9 contiguous amino acids. Accordingly, a fragment can comprise at least about 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50 or more contiguous amino acids. Fragments can retain one or more of the biological activities of the protein, for example as discussed above, as well as fragments that can be used as an immunogen to generate 27875 metalloproteinase antibodies.

Biologically active fragments (peptides which are, for example, 5, 7, 10, 12, 15, 20, 30, 35, 36, 37, 38, 39, 40, 50, 100 or more amino acids in length) can comprise a functional site. Such sites include but are not limited to those discussed above, such as a catalytic site, regulatory site, site important for substrate recognition or binding, zinc binding region, regions containing a metalloproteinase, disintegrin or TSP motif, phosphorylation sites, glycosylation sites, and other functional sites disclosed herein. Such sites or motifs can be identified by means of routine computerized homology searching procedures, such as those disclosed herein.

Fragments, for example, can extend in one or both directions from the functional site to encompass 5, 10, 15, 20, 30, 40, 50, or up to 100 amino acids. Further, fragments can include sub-fragments of the specific sites or regions disclosed herein, which sub-fragments retain the function of the site or region from which they are derived.

The invention also provides fragments with immunogenic properties. These contain an epitope-bearing portion of the 27875 metalloproteinase polypeptide and variants. These epitope-bearing peptides are useful to raise antibodies that bind specifically to an 27875 metalloproteinase polypeptide or region or fragment. These peptides can contain at least 10, 12, at least 14, or between at least about 15 to about 30 amino acids. The epitope-bearing 27875 metalloproteinase polypeptides may be

produced by any conventional means (Houghten, R.A. (1985) *Proc. Natl. Acad. Sci. USA* 82:5131-5135). Simultaneous multiple peptide synthesis is described in U.S. Patent No. 4,631,211.

5 Non-limiting examples of antigenic polypeptides that can be used to generate antibodies include but are not limited to peptides derived from extracellular regions. Regions having a high antigenicity index are shown in Figure 3. However, intracellularly-made antibodies ("intrabodies") are also encompassed, which would recognize intracellular peptide regions.

10 Fragments can be discrete (not fused to other amino acids or polypeptides) or can be within a larger polypeptide. Further, several fragments can be comprised within a single larger polypeptide. In one embodiment a fragment designed for expression in a host can have heterologous pre- and pro-polypeptide regions fused to the amino terminus of the 27875 metalloproteinase polypeptide fragment and an additional region fused to the carboxyl terminus of the fragment.

15 The invention thus provides chimeric or fusion proteins. These comprise an 27875 metalloproteinase peptide sequence operatively linked to a heterologous peptide having an amino acid sequence not substantially homologous to the 27875 metalloproteinase polypeptide. "Operatively linked" indicates that the 27875 metalloproteinase polypeptide and the heterologous peptide are fused in-frame. The
20 heterologous peptide can be fused to the N-terminus or C-terminus of the 27875 metalloproteinase polypeptide or can be internally located.

In one embodiment the fusion protein does not affect 27875 metalloproteinase function *per se*. For example, the fusion protein can be a GST-fusion protein in which 27875 metalloproteinase sequences are fused to the N- or C-terminus of the GST
25 sequences. Other types of fusion proteins include, but are not limited to, enzymatic fusion proteins, for example beta-galactosidase fusions, yeast two-hybrid GAL4 fusions, poly-His fusions and Ig fusions. Such fusion proteins, particularly poly-His fusions, can facilitate the purification of recombinant 27875 metalloproteinase polypeptide. In certain host cells (e.g., mammalian host cells), expression and/or secretion of a protein
30 can be increased by using a heterologous signal sequence. Therefore, in another embodiment, the fusion protein contains a heterologous signal sequence at its C- or N-terminus.

EP-A-O 464 533 discloses fusion proteins comprising various portions of immunoglobulin constant regions. The Fc is useful in therapy and diagnosis and thus results, for example, in improved pharmacokinetic properties (EP-A 0232 262). In drug discovery, for example, human proteins have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists (Bennett *et al.* (1995) *J. Mol. Recog.* 8:52-58 (1995) and Johanson *et al.* *J. Biol. Chem.* 270:9459-9471). Thus, this invention also encompasses soluble fusion proteins containing an 27875 metalloproteinase polypeptide and various portions of the constant regions of heavy or light chains of immunoglobulins of various subclass (IgG, IgM, IgA, IgE). Preferred as immunoglobulin is the constant part of the heavy chain of human IgG, particularly IgG1, where fusion takes place at the hinge region. For some uses it is desirable to remove the Fc after the fusion protein has been used for its intended purpose, for example when the fusion protein is to be used as antigen for immunizations. In a particular embodiment, the Fc part can be removed in a simple way by a cleavage sequence, which is also incorporated and can be cleaved with factor Xa.

A chimeric or fusion protein can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different protein sequences are ligated together in-frame in accordance with conventional techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see Ausubel *et al.* (1992) *Current Protocols in Molecular Biology*). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST protein). An 27875 metalloproteinase -encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to 27875 metalloproteinase.

Another form of fusion protein is one that directly affects 27875 metalloproteinase functions. Accordingly, an 27875 metalloproteinase polypeptide is encompassed by the present invention in which one or more of the 27875 metalloproteinase regions (or parts thereof) has been replaced by heterologous or homologous regions (or parts thereof) from another ADAM-TS or an ADAM. Accordingly, various permutations are possible, for example, as discussed above. Thus,

chimeric 27875 metalloproteinases can be formed in which one or more of the native domains or subregions has been replaced by another. This includes metalloproteinase, disintegrin or thrombospondin domains.

It is understood however that such regions could be derived from an ADAM-TS, ADAM, metalloprotein, disintegrin or thrombospondin that has not yet been characterized. Moreover, disintegrin, metalloprotein, and thrombospondin function can be derived from peptides that contain these functions but are not found in either an ADAM or ADAM-TS family. Accordingly, these domains could be provided from other metalloproteins, disintegrins or thrombospondins.

The isolated 27875 metalloproteinase protein can be purified from cells that naturally express it, such as cells of osteoblast, lung, heart or kidney lineage, especially purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods.

In one embodiment, the protein is produced by recombinant DNA techniques. For example, a nucleic acid molecule encoding the 27875 metalloproteinase polypeptide is cloned into an expression vector, the expression vector introduced into a host cell and the protein expressed in the host cell. The protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques.

Polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally-occurring amino acids. Further, many amino acids, including the terminal amino acids, may be modified by natural processes, such as processing and other post-translational modifications, or by chemical modification techniques well known in the art. Common modifications that occur naturally in polypeptides are described in basic texts, detailed monographs, and the research literature, and they are well known to those of skill in the art.

Accordingly, the polypeptides also encompass derivatives or analogs in which a substituted amino acid residue is not one encoded by the genetic code, in which a substituent group is included, in which the mature polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or in which the additional amino acids are fused to the mature polypeptide, such as a leader or secretory sequence or a sequence for purification of the mature polypeptide or a pro-protein sequence.

Known modifications include, but are not limited to, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent crosslinks, formation of cystine, formation of pyroglutamate, formylation, gamma carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Such modifications are well-known to those of skill in the art and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as *Proteins - Structure and Molecular Properties*, 2nd ed., T.E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as by Wold, F., *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, New York 1-12 (1983); Seifter *et al.* (1990) *Meth. Enzymol.* 182: 626-646) and Rattan *et al.* (1992) *Ann. N.Y. Acad. Sci.* 663:48-62).

As is also well known, polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of post-translation events, including natural processing events and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translational natural processes and by synthetic methods.

Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. Blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally-occurring and synthetic polypeptides. For instance, the aminoterminal residue of polypeptides made in *E. coli*, prior to proteolytic processing, almost invariably will be N-formylmethionine.

The modifications can be a function of how the protein is made. For recombinant polypeptides, for example, the modifications will be determined by the host cell posttranslational modification capacity and the modification signals in the polypeptide amino acid sequence. Accordingly, when glycosylation is desired, a polypeptide should be expressed in a glycosylating host, generally a eukaryotic cell. Insect cells often carry out the same posttranslational glycosylations as mammalian cells and, for this reason, insect cell expression systems have been developed to efficiently express mammalian proteins having native patterns of glycosylation. Similar considerations apply to other modifications.

The same type of modification may be present in the same or varying degree at several sites in a given polypeptide. Also, a given polypeptide may contain more than one type of modification.

Polypeptide Uses

The protein sequences of the present invention can be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to the proteins of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.*, (1997) *Nucleic Acids Res.* 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See <http://www.ncbi.nlm.nih.gov>.

27875 metalloproteinase polypeptides are useful for producing antibodies specific for 27875 metalloproteinase, regions, or fragments. Regions having a high antigenicity index score are shown in Figure 3.

27875 metalloproteinase polypeptides are useful for biological assays related to metalloproteinases, disintegrins or thrombospondins, particularly those functions found in ADAMs and ADAM-TSs. Such assays involve any of the known ADAM, ADAM-

TS, metalloproteinase, disintegrin or thrombospondin functions or activities or properties useful for diagnosis and treatment of 27875 metalloproteinase-related conditions.

These assays include, but are not limited to, binding extracellular matrix, binding integrin, binding zinc or other metals, binding α_2 -macroglobulin, cleaving
5 specific peptide substrates to produce fragments, affecting cell adhesion, binding heparin or other sulfated glycosaminoglycan, such as heparan sulfate, suppressing vascularization, suppressing vascular endothelial growth, breaking down cartilage, inducing apoptosis of endothelial cells, suppressing tumor growth, inhibiting
10 angiogenesis, affecting cellular chemotaxis, affecting cell-cell interaction or cell-matrix interaction, binding integrin, and any of the other biological or functional properties of these proteins, including, but not limited to, those disclosed herein, and in the references cited herein which are incorporated herein by reference for the disclosure of these properties and for the assays based on these properties. Further, assays may relate to changes in the protein, *per se*, and on the effects of these changes, for example, cleavage
15 of the propeptide by furin or other specific proteinase, activation of the protein following cleavage, induction of expression of the protein *in vivo* by LPS, inhibition of function by such agents as SF775, as well as any other effects on the protein mentioned herein or cited in the references herein, which are incorporated herein by reference for these effects and for the subsequent biological consequences of these effects.

20 Such assays include, but are not limited to, those disclosed in Tang *et al.* (*FEBS Letters* 445:223-225 (1999)) (for example, induction by interleukin I *in vitro* and by intravenous administration of lipopolysaccharide *in vivo*, as well as effects on cell adhesion, motility, and growth); Abbaszade *et al.*, above (for example, products resulting from cleavage at the Glu-Ala site in cartilage explants and chondrocyte cultures treated
25 with interleukin I and retinoic acid, determination of aggrecan cleaving activity with and without hydroxamate inhibitors); Kuno *et al.* (1998), above (binding to the extracellular matrix, binding to sulfated glycosaminoglycans, binding to heparan sulfate); Kuno *et al.* (1999) proteinase trapping of α_2 -macroglobulin, furin processing); Tortorella *et al.* (1999), above (detection of aggrecan fragments, especially by neoepitope antibodies,
30 inhibition of cleavage by ADAM-TS inhibitors, inhibition of pro-MMP activation); Vasquez *et al.*, above (suppression of fibroblast growth factor-2-induced vascularization in the cornea pocket assay and inhibition of vascular endothelial growth factor-induced angiogenesis in the chorioallantoic membrane assay, inhibition

of endothelial cell proliferation, competitive inhibition with endostatin, proliferation of human dermal endothelial cells, use of the antiangiogenic region of the TSP-1 motif as bait); Kuno *et al.* (1997), above (heparin binding, induction of expression *in vitro* by interleukin I, induction of expression *in vivo* by LPS); Wolfsberg *et al.*,
5 above (degradation of basement membrane, binding of integrin, and fusogenic activity); Guilpin *et al.* (1988) *J. Biol. Chem.* 273:157-166 (α_2 -macroglobulin trapping, cleavage of prodomain at the furin site to generate active metalloproteinase); Rosendahl *et al.*, above (*J. Biol. Chem.* 272:24588-24593 (1997)) (TNF α processing); Wolfsberg *et al.*, *Developmental Biology* 169:378-383 (1995) (adhesion
10 by integrin binding in the disintegrin domain, antiadhesive function by zinc-dependent metalloproteinase domain). These references are incorporated herein by reference for these specific assays.

Recombinant assay systems include, but are not limited to, those shown in Abbaszade *et al.*, above; Kuno *et al.* (1998), above; Kuno *et al.* (1999), above;
15 Tortorella *et al.*, above; Vasquez *et al.*, above, Kuno *et al.* (1997), above; Wolfsberg *et al.* (*Developmental Biology*), above. These references are also incorporated herein by reference for the cloning and expression systems disclosed therein.

27875 metalloproteinase polypeptides are also useful in drug screening assays, in cell-based or cell-free systems. Cell-based systems can be native, i.e., cells that normally
20 express 27875 metalloproteinase, such as lung, fetal kidney, fetal heart, adult lung and osteoblasts, as a biopsy, or expanded in cell culture. In one embodiment, however, cell-based assays involve recombinant host cells expressing 27875 metalloproteinase. Accordingly, these drug-screening assays can be based on effects on protein function as described above for biological assays useful for diagnosis and treatment.

25 Determining the ability of the test compound to interact with 27875 metalloproteinase can also comprise determining the ability of the test compound to preferentially bind to the polypeptide as compared to the ability of a known binding molecule to bind to the polypeptide.

The polypeptides can be used to identify compounds that modulate 27875
30 metalloproteinase activity. Such compounds, for example, can increase or decrease affinity or rate of binding to substrate, compete with substrate for binding to 27875 metalloproteinase, or displace substrate bound to 27875 metalloproteinase. Both 27875 metalloproteinase and appropriate variants and fragments can be used in high-throughput

screens to assay candidate compounds for the ability to bind to 27875 metalloproteinase. These compounds can be further screened against a functional 27875 metalloproteinase to determine the effect of the compound on 27875 metalloproteinase activity.

Compounds can be identified that activate (agonist) or inactivate (antagonist) 27875 metalloproteinase to a desired degree. Modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject).

27875 metalloproteinase polypeptides can be used to screen a compound for the ability to stimulate or inhibit interaction between 27875 metalloproteinase protein and a target molecule that normally interacts with 27875 metalloproteinase, for example, furin, zinc or other metal, substrate peptide of the metalloproteinase module, substrate of the disintegrin module, for example, integrin, or substrate of the thrombospondin module, i.e., sulfated glycosaminoglycan, such as heparin and heparan sulfate, and accordingly, extracellular matrix. The assay includes the steps of combining 27875 metalloproteinase protein with a candidate compound under conditions that allow the 27875 metalloproteinase protein or fragment to interact with the target molecule, and to detect the formation of a complex between the 27875 metalloproteinase protein and the target or to detect the biochemical consequence of the interaction with 27875 metalloproteinase and the target.

Determining the ability of 27875 metalloproteinase to bind to a target molecule can also be accomplished using a technology such as real-time Bimolecular Interaction Analysis (BIA). Sjolander *et al.* (1991) *Anal. Chem.* 63:2338-2345 and Szabo *et al.* (1995) *Curr. Opin. Struct. Biol.* 5:699-705. As used herein, "BIA" is a technology for studying biospecific interactions in real time, without labeling any of the interactants (e.g., BIAcore™). Changes in the optical phenomenon surface plasmon resonance (SPR) can be used as an indication of real-time reactions between biological molecules.

The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to polypeptide

310); (Ladner *supra*).

Candidate compounds include, for example, 1) peptides such as soluble peptides, including Ig-tailed fusion peptides and members of random peptide libraries (see, e.g., Lam *et al.* (1991) *Nature* 354:82-84; Houghten *et al.* (1991) *Nature* 354:84-86) and
20 combinatorial chemistry-derived molecular libraries made of D- and/or L- configuration amino acids; 2) phosphopeptides (e.g., members of random and partially degenerate, directed phosphopeptide libraries, see, e.g., Songyang *et al.* (1993) *Cell* 72:767-778); 3) antibodies (e.g., polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, and single chain antibodies as well as Fab, F(ab')₂, Fab expression library fragments, and epitope-
25 binding fragments of antibodies); and 4) small organic and inorganic molecules (e.g., molecules obtained from combinatorial and natural product libraries).

One candidate compound is a soluble full-length 27875 metalloproteinase or fragment that competes for peptide, integrin, metal, or glycan binding. Other candidate compounds include mutant 27875 metalloproteinases or appropriate fragments
30 containing mutations that affect 27875 metalloproteinase function and compete for peptide, integrin, metal, or glycan substrate. Accordingly, a fragment that competes for substrate, for example with a higher affinity, or a fragment that binds substrate but does not process or otherwise affect it, is encompassed by the invention.

The invention provides other end points to identify compounds that modulate (stimulate or inhibit) 27875 metalloproteinase activity. The assays typically involve an assay of cellular events that indicate 27875 metalloproteinase activity. Thus, the expression of genes that are up- or down-regulated in response to 27875 metalloproteinase activity can be assayed. In one embodiment, the regulatory region of such genes can be operably linked to a marker that is easily detectable, such as luciferase. Alternatively, modification of 27875 metalloproteinase could also be measured.

Any of the biological or biochemical functions mediated by the 27875 metalloproteinase can be used as an endpoint assay. These include all of the biochemical or biochemical/biological events described herein, in the references cited herein, incorporated by reference for these endpoint assay targets, and other functions known to those of ordinary skill in the art. In the case of the 27875 metalloproteinase, specific end points can include, but are not limited to, the events resulting from expression (or lack thereof) of metalloproteinase, disintegrin or thrombospondin activity. With respect to disorders, this would include, but not be limited to, cartilage breakdown, effects on angiogenesis, such as inhibition, induction of apoptosis of endothelial cells, cell-cell adhesion, as well as cell-matrix interaction stimulation of cell surface receptors by cleavage of extracellular ligand, and resulting clinical effects, such as arthritis and tumor growth. In addition, osteoblast function, differentiation, and proliferation can be assayed as well as the biological effects of osteoblast function such as osteoporosis and osteopetrosis and other disorders and pathology, such as that disclosed above, for bone-forming cells.

Binding and/or activating compounds can also be screened by using chimeric 27875 metalloproteinase proteins in which one or more regions, segments, sites, and the like, as disclosed herein, or parts thereof, can be replaced by heterologous and homologous counterparts derived from other ADAM-TSs, ADAMs, metalloproteinases, disintegrins or thrombospondins. For example, a catalytic region can be used that interacts with a different peptide or glycan specificity and/or affinity than the native 27875 metalloproteinase. Accordingly, a different set of components is available as an end-point assay for activation. As a further alternative, the site of modification by an effector protein, for example phosphorylation, can be replaced with the site for a different effector protein. Activation can also be detected by a reporter gene containing

Thus, the soluble polypeptide that competes with the target 27875 metalloproteinase region is designed to contain peptide sequences corresponding to the region of interest.

15 Another type of competition-binding assay can be used to discover compounds that interact with specific functional sites. As an example, bindable zinc and a candidate compound can be added to a sample of 27875 metalloproteinase. Compounds that interact with 27875 metalloproteinase at the same site as the zinc will reduce the amount of complex formed between 27875 metalloproteinase and the zinc. Accordingly, it is
20 possible to discover a compound that specifically prevents interaction between 27875 metalloproteinase and the zinc component. Another example involves adding a candidate compound to a sample of 27875 metalloproteinase and substrate peptide. A compound that competes with the peptide will reduce the amount of hydrolysis or binding of the peptide to 27875 metalloproteinase. Accordingly, compounds can be
25 discovered that directly interact with 27875 metalloproteinase and compete with the peptide. Such assays can involve any other component that interacts with 27875 metalloproteinase, such as integrin or sulfated glycosaminoglycan.

 To perform cell free drug screening assays, it is desirable to immobilize either 27875 metalloproteinase, or fragment, or its target molecule to facilitate separation of
30 complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay.

 Techniques for immobilizing proteins on matrices can be used in the drug screening assays. In one embodiment, a fusion protein can be provided which adds a

domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase/27875 metalloproteinase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the cell lysates (e.g., ³⁵S-labeled) and the candidate
5 compound, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly, or in the supernatant after the complexes is dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-
10 PAGE, and the level of 27875 metalloproteinase-binding protein found in the bead fraction quantitated from the gel using standard electrophoretic techniques. For example, either the polypeptide or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin using techniques well known in the art. Alternatively, antibodies reactive with the protein but which do not interfere with
15 binding of the protein to its target molecule can be derivatized to the wells of the plate, and the protein trapped in the wells by antibody conjugation. Preparations of an 27875 metalloproteinase-binding target component, such as a peptide or zinc component, and a candidate compound are incubated in 27875 metalloproteinase-presenting wells and the amount of complex trapped in the well can be quantitated. Methods for detecting such
20 complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with 27875 metalloproteinase target molecule, or which are reactive with 27875 metalloproteinase and compete with the target molecule; as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the target molecule.
25 Modulators of 27875 metalloproteinase activity identified according to these drug screening assays can be used to treat a subject with a disorder related to 27875 metalloproteinase, by treating cells that express the 27875 metalloproteinase. These methods of treatment include the steps of administering the modulators of 27875 metalloproteinase activity in a pharmaceutical composition as described herein, to a
30 subject in need of such treatment.

27875 metalloproteinase is highly expressed in fetal kidney, fetal heart, and undifferentiated osteoblasts. As such it is specifically involved in disorders relating to these tissues. Examples include, but are not limited to, osteoporosis and

osteopetrosis, as well as other disorders involving osteoblast differentiation, function, and growth. Furthermore, expression is also relevant to disorders of several other tissues, as shown in Figures 5-7. Disorders of these tissues are disclosed hereinabove. 27875 metalloproteinase polypeptides are thus useful for treating an 27875

5 metalloproteinase-associated disorder characterized by aberrant expression or activity of an 27875 metalloproteinase. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., upregulates or downregulates) expression or activity of the protein. In another embodiment, the method involves
10 administering 27875 metalloproteinase as therapy to compensate for reduced or aberrant expression or activity of the protein.

Methods for treatment include but are not limited to the use of soluble 27875 metalloproteinase or fragments of 27875 metalloproteinase protein that compete for substrate or any other component that directly interacts with 27875 metalloproteinase,
15 such as integrin, glycan, zinc, or any of the enzymes that modify 27875 metalloproteinase. These 27875 metalloproteinases or fragments can have a higher affinity for the target so as to provide effective competition.

Stimulation of activity is desirable in situations in which the protein is abnormally downregulated and/or in which increased activity is likely to have a
20 beneficial effect. Likewise, inhibition of activity is desirable in situations in which the protein is abnormally upregulated and/or in which decreased activity is likely to have a beneficial effect. In one example of such a situation, a subject has a disorder characterized by aberrant development or cellular differentiation. In another example, the subject has a disorder characterized by an aberrant hematopoietic response. In
25 another example, it is desirable to achieve tissue regeneration in a subject (e.g., where a subject has undergone bone trauma or osteoporosis).

In yet another aspect of the invention, the proteins of the invention can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J. Biol.*
30 *Chem.* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO 94/10300), to identify other proteins (captured proteins) which bind to or interact with the proteins of the invention and modulate their activity.

27875 metalloproteinase polypeptides also are useful to provide a target for diagnosing a disease or predisposition to disease mediated by 27875 metalloproteinase, including, but not limited to, those diseases discussed herein, and particularly bone-related disorders, as disclosed above. Targets are useful for diagnosing a disease or predisposition to disease mediated by 27875 metalloproteinase. Accordingly, methods are provided for detecting the presence, or levels of, 27875 metalloproteinase in a cell, tissue, or organism. The method involves contacting a biological sample with a compound capable of interacting with 27875 metalloproteinase such that the interaction can be detected. One agent for detecting 27875 metalloproteinase is an antibody capable of selectively binding to 27875 metalloproteinase. A biological sample includes tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject.

The 27875 metalloproteinase also provides a target for diagnosing active disease, or predisposition to disease, in a patient having a variant 27875 metalloproteinase. Thus, 27875 metalloproteinase can be isolated from a biological sample and assayed for the presence of a genetic mutation that results in an aberrant protein. This includes amino acid substitution, deletion, insertion, rearrangement, (as the result of aberrant splicing events), and inappropriate post-translational modification. Analytic methods include altered electrophoretic mobility, altered tryptic peptide digest, altered 27875 metalloproteinase activity in cell-based or cell-free assay, alteration in peptide binding or degradation, integrin binding, glycan binding, zinc binding or antibody-binding pattern, altered isoelectric point, direct amino acid sequencing, and any other of the known assay techniques useful for detecting mutations in a protein in general or in an 27875 metalloproteinase specifically, such as are disclosed herein.

In vitro techniques for detection of 27875 metalloproteinase include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. Alternatively, the protein can be detected *in vivo* in a subject by introducing into the subject a labeled anti-27875 metalloproteinase antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques. Particularly useful are methods, which detect the allelic variant of 27875 metalloproteinase expressed in a subject, and methods, which detect fragments of 27875 metalloproteinase in a sample.

of drug action. Thus, the pharmacogenomes of the individual permit the selection of effective compounds and effective dosages of such compounds for prophylactic or therapeutic treatment based on the individual's genotype. The discovery of genetic polymorphisms in some drug metabolizing enzymes has explained why some patients do
15 not obtain the expected drug effects, show an exaggerated drug effect, or experience serious toxicity from standard drug dosages. Polymorphisms can be expressed in the phenotype of the extensive metabolizer and the phenotype of the poor metabolizer. Accordingly, genetic polymorphism may lead to allelic protein variants of 27875 metalloproteinase in which one or more of 27875 metalloproteinase functions in one
20 population is different from those in another population. The polypeptides thus allow a target to ascertain a genetic predisposition that can affect treatment modality. Thus, in a peptide-based treatment, polymorphism may give rise to catalytic regions that are more or less active. Accordingly, dosage would necessarily be modified to maximize the therapeutic effect within a given population containing the polymorphism. As an
25 alternative to genotyping, specific polymorphic polypeptides could be identified.

27875 metalloproteinase polypeptides are also useful for monitoring therapeutic effects during clinical trials and other treatment. Thus, the therapeutic effectiveness of an agent that is designed to increase or decrease gene expression, protein levels or 27875 metalloproteinase activity can be monitored over the course of treatment using 27875
30 metalloproteinase polypeptides as an end-point target. The monitoring can be, for example, as follows: (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression or activity of the protein in the pre-administration sample; (iii) obtaining one or more post-

administration samples from the subject; (iv) detecting the level of expression or activity of the protein in the post-administration samples; (v) comparing the level of expression or activity of the protein in the pre-administration sample with the protein in the post-administration sample or samples; and (vi) increasing or decreasing the
5 administration of the agent to the subject accordingly.

Antibodies

The invention also provides antibodies that selectively bind to 27875 metalloproteinase and its variants and fragments. An antibody is considered to
10 selectively bind, even if it also binds to other proteins that are not substantially homologous with 27875 metalloproteinase. These other proteins share homology with a fragment or domain of 27875 metalloproteinase. This conservation in specific regions gives rise to antibodies that bind to both proteins by virtue of the homologous sequence. In this case, it would be understood that antibody binding to 27875 metalloproteinase is
15 still selective.

Antibodies can be polyclonal or monoclonal. An intact antibody, or a fragment thereof (e.g. Fab or F(ab')₂) can be used. An appropriate immunogenic preparation can be derived from native, recombinantly expressed, or chemically synthesized peptides.

To generate antibodies, an isolated 27875 metalloproteinase polypeptide is used
20 as an immunogen to generate antibodies using standard techniques for polyclonal and monoclonal antibody preparation. Either the full-length protein or antigenic peptide fragment can be used. Regions having a high antigenicity index are shown in Figure 3.

Antibodies are preferably prepared from these regions or from discrete fragments in these regions. However, antibodies can be prepared from any region of
25 the peptide as described herein. A preferred fragment produces an antibody that diminishes or completely prevents peptide hydrolysis or binding. Antibodies can be developed against the entire 27875 metalloproteinase or domains of 27875 metalloproteinase as described herein, for example, the zinc binding region, metalloproteinase motif, the disintegrin domain, the TSP motif, or subregions thereof.
30 Antibodies can also be developed against specific functional sites as disclosed herein.

The antigenic peptide can comprise a contiguous sequence of at least 12, 14, 15, or 30 amino acid residues. In one embodiment, fragments correspond to regions that are located on the surface of the protein, e.g., hydrophilic regions. These fragments are not

to be construed, however, as encompassing any fragments, which may be disclosed prior to the invention.

Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H .

15 Antibody Uses

The antibodies can be used to isolate a 27875 metalloproteinase by standard techniques, such as affinity chromatography or immunoprecipitation. The antibodies can facilitate the purification of the natural 27875 metalloproteinase from cells and recombinantly produced 27875 metalloproteinase expressed in host cells.

20 The antibodies are useful to detect the presence of 27875 metalloproteinase in cells or tissues to determine the pattern of expression of 27875 metalloproteinase among various tissues in an organism and over the course of normal development. The antibodies can be used to detect 27875 metalloproteinase *in situ*, *in vitro*, or in a cell lysate or supernatant in order to evaluate the abundance and pattern of expression.

25 Antibody detection of circulating fragments of the full length 27875 metalloproteinase can be used to identify 27875 metalloproteinase turnover. In addition, the antibodies can be used to assess abnormal tissue distribution or abnormal expression during development.

Further, the antibodies can be used to assess 27875 metalloproteinase expression in disease states such as in active stages of the disease or in an individual with a predisposition toward disease related to 27875 metalloproteinase function. When a disorder is caused by an inappropriate tissue distribution, developmental expression, or level of expression of 27875 metalloproteinase protein, the antibody can be prepared

against the normal 27875 metalloproteinase protein. If a disorder is characterized by a specific mutation in 27875 metalloproteinase, antibodies specific for this mutant protein can be used to assay for the presence of the specific mutant 27875 metalloproteinase. However, intracellularly-made antibodies ("intrabodies") are also encompassed, which
5 would recognize intracellular 27875 metalloproteinase peptide regions.

The antibodies can also be used to assess normal and aberrant subcellular localization of cells in the various tissues in an organism. Antibodies can be developed against the whole 27875 metalloproteinase or portions of 27875 metalloproteinase.

The diagnostic uses can be applied, not only in genetic testing, but also in
10 monitoring a treatment modality. Accordingly, where treatment is ultimately aimed at correcting 27875 metalloproteinase expression level or the presence of aberrant 27875 metalloproteinases and aberrant tissue distribution or developmental expression, antibodies directed against 27875 metalloproteinase or relevant fragments can be used to monitor therapeutic efficacy.

15 Additionally, antibodies are useful in pharmacogenomic analysis. Thus, antibodies prepared against polymorphic 27875 metalloproteinase can be used to identify individuals that require modified treatment modalities.

The antibodies are also useful as diagnostic tools as an immunological marker for aberrant 27875 metalloproteinase analyzed by electrophoretic mobility, isoelectric
20 point, tryptic peptide digest, and other physical assays known to those in the art.

The antibodies are also useful for tissue typing. Thus, where a specific 27875 metalloproteinase has been correlated with expression in a specific tissue, antibodies that are specific for this 27875 metalloproteinase can be used to identify a tissue type.

The antibodies are also useful in forensic identification. Accordingly, where an
25 individual has been correlated with a specific genetic polymorphism resulting in a specific polymorphic protein, an antibody specific for the polymorphic protein can be used as an aid in identification.

The antibodies are also useful for inhibiting 27875 metalloproteinase function, for example, zinc binding, metalloproteinase activity, disintegrin activity or TSP activity.
30 For example, metalloproteinase activity may be measured by the ability to form a covalent binding complex with α_2 -macroglobulin (Kuno *et al.* (1999) *J Biol Chem* 274:18821-18826).

15 antibodies such as a labeled or labelable antibody and a compound or agent for detecting
27875 metalloproteinase in a biological sample; means for determining the amount of
27875 metalloproteinase in the sample; and means for comparing the amount of 27875
metalloproteinase in the sample with a standard. The compound or agent can be
packaged in a suitable container. The kit can further comprise instructions for using the
20 kit to detect 27875 metalloproteinase.

Polynucleotides

The invention provides isolated polynucleotides encoding the novel 27875
metalloproteinase. The term “27875 metalloproteinase polynucleotide” or “27875
25 metalloproteinase nucleic acid” refers to the sequence shown in SEQ ID NO:2. The
term “27875 metalloproteinase polynucleotide” or “27875 metalloproteinase nucleic
acid” further includes variants and fragments of 27875 metalloproteinase
polynucleotides.

An “isolated” 27875 metalloproteinase nucleic acid is one that is separated from
30 other nucleic acid present in the natural source of 27875 metalloproteinase nucleic acid.
Preferably, an “isolated” nucleic acid is free of sequences which naturally flank 27875
metalloproteinase nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic
acid) in the genomic DNA of the organism from which the nucleic acid is derived.

However, there can be some flanking nucleotide sequences, for example up to about 5KB. The important point is that the 27875 metalloproteinase nucleic acid is isolated from flanking sequences such that it can be subjected to the specific manipulations described herein, such as recombinant expression, preparation of probes and primers, and other uses specific to the 27875 metalloproteinase nucleic acid sequences. In one embodiment, the 27875 metalloproteinase nucleic acid comprises only the coding region.

Moreover, an "isolated" nucleic acid molecule, such as a cDNA or RNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. However, the nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated.

In some instances, the isolated material will form part of a composition (for example, a crude extract containing other substances), buffer system or reagent mix. In other circumstances, the material may be purified to essential homogeneity, for example as determined by PAGE or column chromatography such as HPLC. Preferably, an isolated nucleic acid comprises at least about 50, 80 or 90% (on a molar basis) of all macromolecular species present.

For example, recombinant DNA molecules contained in a vector are considered isolated. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the isolated DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

In some instances, the isolated material will form part of a composition (or example, a crude extract containing other substances), buffer system or reagent mix. In other circumstances, the material may be purified to essential homogeneity, for example as determined by PAGE or column chromatography such as HPLC. Preferably, an isolated nucleic acid comprises at least about 50, 80 or 90% (on a molar basis) of all macromolecular species present.

27875 metalloproteinase polynucleotides can encode the mature protein plus additional amino or carboxyterminal amino acids, or amino acids interior to the mature

polypeptide (when the mature form has more than one polypeptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, facilitate protein trafficking, prolong or shorten protein half-life or facilitate manipulation of a protein for assay or production, among other things. As generally is
5 the case *in situ*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

27875 metalloproteinase polynucleotides include, but are not limited to, the sequence encoding the mature polypeptide alone, the sequence encoding the mature polypeptide and additional coding sequences, such as a leader or secretory sequence
10 (e.g., a pre-pro or pro-protein sequence), the sequence encoding the mature polypeptide, with or without the additional coding sequences, plus additional non-coding sequences, for example introns and non-coding 5' and 3' sequences such as transcribed but non-translated sequences that play a role in transcription, mRNA processing (including splicing and polyadenylation signals), ribosome binding and stability of mRNA. In
15 addition, the polynucleotide may be fused to a marker sequence encoding, for example, a peptide that facilitates purification.

27875 metalloproteinase polynucleotides can be in the form of RNA, such as mRNA, or in the form DNA, including cDNA and genomic DNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The nucleic
20 acid, especially DNA, can be double-stranded or single-stranded. Single-stranded nucleic acid can be the coding strand (sense strand) or the non-coding strand (anti-sense strand).

The invention further provides variant 27875 metalloproteinase polynucleotides, and fragments thereof, that differ from the nucleotide sequence shown in SEQ ID NO:2
25 due to degeneracy of the genetic code and thus encode the same protein as that encoded by the nucleotide sequence shown in SEQ ID NO:2.

The invention also provides 27875 metalloproteinase nucleic acid molecules encoding the variant polypeptides described herein. Such polynucleotides may be naturally occurring, such as allelic variants (same locus), homologs (different locus), and
30 orthologs (different organism), or may be constructed by recombinant DNA methods or by chemical synthesis. Such non-naturally occurring variants may be made by mutagenesis techniques, including those applied to polynucleotides, cells, or organisms.

Accordingly, as discussed above, the variants can contain nucleotide substitutions, deletions, inversions and insertions.

Typically, variants have a substantial identity with a nucleic acid molecules of SEQ ID NO:2 and the complements thereof. Variation can occur in either or both the
5 coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions.

Orthologs, homologs, and allelic variants can be identified using methods well known in the art. These variants comprise a nucleotide sequence encoding a 27875 metalloproteinase that is typically at least about 60-65%, 65-70%, 70-75%, more
10 typically at least about 80-85%, and most typically at least about 90-95% or more homologous to the nucleotide sequence shown in SEQ ID NO:2 or a fragment of this sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under stringent conditions, to the nucleotide sequence shown in SEQ ID NO:2 or a fragment of the sequence. It is understood that stringent hybridization does not
15 indicate substantial homology where it is due to general homology, such as polyA⁺ sequences, or sequences common to all or most proteins, metalloproteinases, zinc binding proteins, thrombospondins, disintegrins, ADAMs, proteins in the ADAM-TS family, or even all proteins in specific ADAM-TS subfamilies, such as ADAM-TS-1, 3, etc. Moreover, it is understood that variants do not include any of the nucleic acid
20 sequences that may have been disclosed prior to the invention.

As used herein, the term "stringent conditions" is intended to describe conditions comprising hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC/0.1% SDS at 65°C. Methods of hybridization are known to those skilled in the art and can be found in *Current Protocols*
25 *in Molecular Biology*, John Wiley & Sons, N.Y. (1998), incorporated by reference. In one embodiment, an isolated nucleic acid molecule that hybridizes under stringent conditions to the sequence of SEQ ID NO:2 corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g.,
30 encodes a natural protein).

The present invention also provides isolated nucleic acids that contain a single or double stranded fragment or portion that hybridizes under stringent conditions to the nucleotide sequence of SEQ ID NO:2 or the complement of SEQ ID NO:2. In one

embodiment, the nucleic acid consists of a portion of the nucleotide sequence of SEQ ID NO:2 and the complement of SEQ ID NO:2. The nucleic acid fragments of the invention are at least about 15, preferably at least about 16, 17, 18, 19, 20, 23 or 25 contiguous nucleotides, and can be 30, 33, 35, 40, 50, 60, 70, 75, 80, 90, 100, 200, 5 500 or more nucleotides in length. Longer fragments, for example, 600 or more nucleotides in length, which encode antigenic proteins or polypeptides described herein are also useful.

Furthermore, the invention provides polynucleotides that comprise a fragment of the full-length 27875 metalloproteinase polynucleotides. The fragment can be single 10 or double-stranded and can comprise DNA or RNA. The fragment can be derived from either the coding or the non-coding sequence.

In one embodiment, the nucleic acid sequence is selected from the group consisting of:

(a) a nucleotide sequence encoding a fragment of the amino acid sequence 15 shown in SEQ ID NO:1, wherein the fragment comprises at least 26 contiguous amino acids;

(b) a nucleotide sequence comprising at least 75 consecutive nucleotides of the sequence shown in SEQ ID NO:2;

(c) a nucleotide sequence comprising at least 33 consecutive nucleotides of 20 residues 1-4800 of SEQ ID NO:2;

(d) a nucleotide sequence encoding residues 31-1687 of the amino acid shown in SEQ ID NO:1;

(e) a nucleotide sequence encoding residues 244-259 of SEQ ID NO:1;

(f) a nucleotide sequence encoding residues 385-394 of SEQ ID NO:1;

25 (g) a nucleotide sequence encoding residues 541-592 of SEQ ID NO:1;

(h) a nucleotide sequence encoding residues 542-592 of SEQ ID NO:1;

(i) a nucleotide sequence encoding residues 825-868 of SEQ ID NO:1;

(j) a nucleotide sequence encoding residues 949-988 of SEQ ID NO:1;

(k) a nucleotide sequence encoding residues 1415-1463 of SEQ ID NO:1;

30 and

(l) a nucleotide sequence complementary to a nucleotide sequences of (a)-(k).

In another embodiment an isolated 27875 metalloproteinase nucleic acid encodes the entire coding region. In another embodiment the isolated 27875 metalloproteinase nucleic acid encodes a sequence corresponding to the mature protein that may be from about amino acid 6 to the last amino acid. Other fragments include nucleotide sequences encoding the amino acid fragments described herein.

Thus, 27875 metalloproteinase nucleic acid fragments further include sequences corresponding to the regions described herein, subregions also described, and specific functional sites. 27875 metalloproteinase nucleic acid fragments also include combinations of the regions, segments, motifs, and other functional sites described above. It is understood that a 27875 metalloproteinase fragment includes any nucleic acid sequence that does not include the entire gene. A person of ordinary skill in the art would be aware of the many permutations that are possible. Nucleic acid fragments, according to the present invention, are not to be construed as encompassing those fragments that may have been disclosed prior to the invention.

Where the location of the regions or sites have been predicted by computer analysis, one of ordinary skill would appreciate that the amino acid residues constituting these regions can vary depending on the criteria used to define the regions.

Polynucleotide Uses

The nucleotide sequences of the present invention can be used as a "query sequence" to perform a search against public databases, for example, to identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-10. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to the proteins of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.* (1997) *Nucleic Acids Res.* 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See <http://www.ncbi.nlm.nih.gov>.

The nucleic acid fragments of the invention provide probes or primers in assays such as those described below. "Probes" are oligonucleotides that hybridize in a base-specific manner to a complementary strand of nucleic acid. Such probes

include polypeptide nucleic acids, as described in Nielsen *et al.* (1991) *Science* 254:1497-1500. Typically, a probe comprises a region of nucleotide sequence that hybridizes under highly stringent conditions to at least about 15, typically about 20-25, and more typically about 40, 50 or 75 consecutive nucleotides of the nucleic acid sequence shown in SEQ ID NO:2 and the complements thereof. More typically, the probe further comprises a label, e.g., radioisotope, fluorescent compound, enzyme, or enzyme co-factor.

As used herein, the term "primer" refers to a single-stranded oligonucleotide which acts as a point of initiation of template-directed DNA synthesis using well-known methods (e.g., PCR, LCR) including, but not limited to those described herein. The appropriate length of the primer depends on the particular use, but typically ranges from about 15 to 30 nucleotides. The term "primer site" refers to the area of the target DNA to which a primer hybridizes. The term "primer pair" refers to a set of primers including a 5' (upstream) primer that hybridizes with the 5' end of the nucleic acid sequence to be amplified and a 3' (downstream) primer that hybridizes with the complement of the sequence to be amplified.

27875 metalloproteinase polynucleotides are thus useful for probes, primers, and in biological assays. Where the polynucleotides are used to assess 27875 metalloproteinase properties or functions, such as in the assays described herein, all or less than all of the entire cDNA can be useful. Assays specifically directed to 27875 metalloproteinase functions, such as assessing agonist or antagonist activity, encompass the use of known fragments. Further, diagnostic methods for assessing 27875 metalloproteinase function can also be practiced with any fragment, including those fragments that may have been known prior to the invention. Similarly, in methods involving treatment of 27875 metalloproteinase dysfunction, all fragments are encompassed including those, which may have been known in the art.

27875 metalloproteinase polynucleotides are useful as a hybridization probe for cDNA and genomic DNA to isolate a full-length cDNA and genomic clones encoding the polypeptides described in SEQ ID NO:1 and to isolate cDNA and genomic clones that correspond to variants producing the same polypeptides shown in SEQ ID NO:1 or the other variants described herein. Variants can be isolated from the same tissue and organism from which the polypeptides shown in SEQ ID NO:1 were isolated, different tissues from the same organism, or from different organisms. This method is useful for

isolating genes and cDNA that are developmentally-controlled and therefore may be expressed in the same tissue or different tissues at different points in the development of an organism.

The probe can correspond to any sequence along the entire length of the gene
5 encoding the 27875 metalloproteinase polypeptide. Accordingly, it could be derived from 5' noncoding regions, the coding region, and 3' noncoding regions.

The nucleic acid probe can be, for example, the full-length cDNA of SEQ ID NO:2, or a fragment thereof, such as an oligonucleotide of at least 12, 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent
10 conditions to mRNA or DNA.

Fragments of the polynucleotides described herein are also useful to synthesize larger fragments or full-length polynucleotides described herein, ribozymes or antisense molecules. For example, a fragment can be hybridized to any portion of an mRNA and a larger or full-length cDNA can be produced.

15 Antisense nucleic acids of the invention can be designed using the nucleotide sequences of SEQ ID NO:2, and constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase
20 the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-
25 methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-
30

15 PNAs has been shown to allow for specific hybridization to DNA and RNA under
conditions of low ionic strength. The synthesis of PNA oligomers can be performed
using standard solid phase peptide synthesis protocols as described in Hyrup *et al.*
(1996), *supra*; Perry-O'Keefe *et al.* (1996) *Proc. Natl. Acad. Sci. USA* 93:14670.
PNAs can be further modified, e.g., to enhance their stability, specificity or cellular
20 uptake, by attaching lipophilic or other helper groups to PNA, by the formation of
PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery
known in the art. The synthesis of PNA-DNA chimeras can be performed as
described in Hyrup (1996), *supra*, Finn *et al.* (1996) *Nucleic Acids Res.* 24(17):3357-
63, Mag *et al.* (1989) *Nucleic Acids Res.* 17:5973, and Peterser *et al.* (1975)
25 *Bioorganic Med. Chem. Lett.* 5:1119.

The nucleic acid molecules and fragments of the invention can also include
other appended groups such as peptides (e.g., for targeting host cell 27875
metalloproteinases *in vivo*), or agents facilitating transport across the cell membrane
(see, e.g., Letsinger *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre
30 *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. WO
88/0918) or the blood brain barrier (see, e.g., PCT Publication No. WO 89/10134). In
addition, oligonucleotides can be modified with hybridization-triggered cleavage

agents (see, e.g., Krol *et al.* (1988) *Bio-Techniques* 6:958-976) or intercalating agents (see, e.g., Zon (1988) *Pharm Res.* 5:539-549).

27875 metalloproteinase polynucleotides are also useful as primers for PCR to amplify any given region of an 27875 metalloproteinase polynucleotide.

- 5 27875 metalloproteinase polynucleotides are also useful for constructing recombinant vectors. Such vectors include expression vectors that express a portion of, or all of, the 27875 metalloproteinase polypeptides. Vectors also include insertion vectors, used to integrate into another polynucleotide sequence, such as into the cellular genome, to alter *in situ* expression of 27875 metalloproteinase genes and gene products.
- 10 For example, an endogenous 27875 metalloproteinase coding sequence can be replaced via homologous recombination with all or part of the coding region containing one or more specifically introduced mutations.

27875 metalloproteinase polynucleotides are also useful for expressing antigenic portions of 27875 metalloproteinase proteins.

- 15 27875 metalloproteinase polynucleotides are also useful as probes for determining the chromosomal positions of 27875 metalloproteinase polynucleotides by means of *in situ* hybridization methods, such as FISH. (For a review of this technique, see Verma *et al.* (1988) *Human Chromosomes: A Manual of Basic Techniques* (Pergamon Press, New York), and PCR mapping of somatic cell hybrids. The mapping
- 20 of the sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

- Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to
- 25 noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

- Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map
- 30 data. (Such data are found, for example, in V. McKusick, *Mendelian Inheritance in Man*, available on-line through Johns Hopkins University Welch Medical Library). The relationship between a gene and a disease mapped to the same chromosomal region, can

occurred and whether the duplication occurs in all or only a subset of tissues. The genes can be naturally occurring or can have been introduced into a cell, tissue, or organism exogenously.

27875 metalloproteinase polynucleotides are also useful for designing ribozymes
20 corresponding to all, or a part, of the mRNA produced from genes encoding the polynucleotides described herein.

27875 metalloproteinase polynucleotides are also useful for constructing host cells expressing a part, or all, of 27875 metalloproteinase polynucleotides and polypeptides.

25 27875 metalloproteinase polynucleotides are also useful for constructing transgenic animals expressing all, or a part, of 27875 metalloproteinase polynucleotides and polypeptides.

27875 metalloproteinase polynucleotides are also useful for making vectors that express part, or all, of 27875 metalloproteinase polypeptides.

30 27875 metalloproteinase polynucleotides are also useful as hybridization probes for determining the level of 27875 metalloproteinase nucleic acid expression.

Accordingly, the probes can be used to detect the presence of, or to determine levels of, 27875 metalloproteinase nucleic acid in cells, tissues, and in organisms. The nucleic

acid whose level is determined can be DNA or RNA. Accordingly, probes corresponding to the polypeptides described herein can be used to assess gene copy number in a given cell, tissue, or organism. This is particularly relevant in cases in which there has been an amplification of 27875 metalloproteinase genes.

5 Alternatively, the probe can be used in an *in situ* hybridization context to assess the position of extra copies of 27875 metalloproteinase genes, as on extrachromosomal elements or as integrated into chromosomes in which the 27875 metalloproteinase gene is not normally found, for example as a homogeneously staining region.

These uses are relevant for diagnosis of disorders involving an increase or
10 decrease in 27875 metalloproteinase expression relative to normal, such as a proliferative disorder, a differentiative or developmental disorder, or a hematopoietic disorder. Disorders in which 27875 metalloproteinase expression is relevant include, but are not limited to, those of bone, such as osteoporosis and osteopetrosis.

Thus, the present invention provides a method for identifying a disease or
15 disorder associated with aberrant expression or activity of 27875 metalloproteinase nucleic acid, in which a test sample is obtained from a subject and nucleic acid (e.g., mRNA, genomic DNA) is detected, wherein the presence of the nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant expression or activity of the nucleic acid.

20 One aspect of the invention relates to diagnostic assays for determining nucleic acid expression as well as activity in the context of a biological sample (e.g., blood, serum, cells, tissue) to determine whether an individual has a disease or disorder, or is at risk of developing a disease or disorder, associated with aberrant nucleic acid expression or activity. Such assays can be used for prognostic or
25 predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with expression or activity of the nucleic acid molecules.

In vitro techniques for detection of mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detecting DNA includes Southern
30 hybridizations and *in situ* hybridization.

Probes can be used as a part of a diagnostic test kit for identifying cells or tissues that express 27875 metalloproteinase, such as by measuring the level of an 27875

metalloproteinase-encoding nucleic acid in a sample of cells from a subject e.g., mRNA or genomic DNA, or determining if the 27875 metalloproteinase gene has been mutated.

Nucleic acid expression assays are useful for drug screening to identify compounds that modulate 27875 metalloproteinase nucleic acid expression (e.g., antisense, polypeptides, peptidomimetics, small molecules or other drugs). A cell is contacted with a candidate compound and the expression of mRNA determined. The level of expression of the mRNA in the presence of the candidate compound is compared to the level of expression of the mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder characterized by aberrant nucleic acid expression. The modulator can bind to the nucleic acid or indirectly modulate expression, such as by interacting with other cellular components that affect nucleic acid expression.

Modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject) in patients or in transgenic animals. The invention thus provides a method for identifying a compound that can be used to treat a disorder associated with nucleic acid expression of the 27875 metalloproteinase gene. The method typically includes assaying the ability of the compound to modulate the expression of the 27875 metalloproteinase nucleic acid and thus identifying a compound that can be used to treat a disorder characterized by undesired 27875 metalloproteinase nucleic acid expression.

The assays can be performed in cell-based and cell-free systems. Cell-based assays include cells naturally expressing the 27875 metalloproteinase nucleic acid or recombinant cells genetically engineered to express specific nucleic acid sequences. Alternatively, candidate compounds can be assayed *in vivo* in patients or in transgenic animals.

The assay for 27875 metalloproteinase nucleic acid expression can involve direct assay of nucleic acid levels, such as mRNA levels, or on collateral compounds (such as peptide hydrolysis). Further, the expression of genes that are up- or down-regulated in response to 27875 metalloproteinase activity can also be assayed. In this embodiment the regulatory regions of these genes can be operably linked to a reporter gene such as luciferase.

Thus, modulators of 27875 metalloproteinase gene expression can be identified in a method wherein a cell is contacted with a candidate compound and the expression of mRNA determined. The level of expression of 27875 metalloproteinase mRNA in the presence of the candidate compound is compared to the level of expression of 27875 metalloproteinase mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder characterized by aberrant nucleic acid expression. When expression of mRNA is statistically significantly greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of nucleic acid expression. When nucleic acid expression is statistically significantly less in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of nucleic acid expression.

Accordingly, the invention provides methods of treatment, with the nucleic acid as a target, using a compound identified through drug screening as a gene modulator to modulate 27875 metalloproteinase nucleic acid expression. Modulation includes both up-regulation (i.e. activation or agonization) or down-regulation (suppression or antagonization) or effects on nucleic acid activity (e.g. when nucleic acid is mutated or improperly modified). Treatment is of disorders characterized by aberrant expression or activity of the nucleic acid.

Alternatively, a modulator for 27875 metalloproteinase nucleic acid expression can be a small molecule or drug identified using the screening assays described herein as long as the drug or small molecule inhibits 27875 metalloproteinase nucleic acid expression.

27875 metalloproteinase polynucleotides are also useful for monitoring the effectiveness of modulating compounds on the expression or activity of the 27875 metalloproteinase gene in clinical trials or in a treatment regimen. Thus, the gene expression pattern can serve as a barometer for the continuing effectiveness of treatment with the compound, particularly with compounds to which a patient can develop resistance. The gene expression pattern can also serve as a marker indicative of a physiological response of the affected cells to the compound. Accordingly, such monitoring would allow either increased administration of the compound or the administration of alternative compounds to which the patient has not become resistant.

Similarly, if the level of nucleic acid expression falls below a desirable level, administration of the compound could be commensurately decreased.

Monitoring can be, for example, as follows: (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of a specified mRNA or genomic DNA of the invention in the pre-administration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the mRNA or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the mRNA or genomic DNA in the pre-administration sample with the mRNA or genomic DNA in the post-administration sample or samples; and (vi) increasing or decreasing the administration of the agent to the subject accordingly.

27875 metalloproteinase polynucleotides are also useful in diagnostic assays for qualitative changes in 27875 metalloproteinase nucleic acid, and particularly in qualitative changes that lead to pathology. The polynucleotides can be used to detect mutations in 27875 metalloproteinase genes and gene expression products such as mRNA. The polynucleotides can be used as hybridization probes to detect naturally-occurring genetic mutations in the 27875 metalloproteinase gene and thereby to determine whether a subject with the mutation is at risk for a disorder caused by the mutation. Mutations include deletion, addition, or substitution of one or more nucleotides in the gene, chromosomal rearrangement, such as inversion or transposition, modification of genomic DNA, such as aberrant methylation patterns or changes in gene copy number, such as amplification. Detection of a mutated form of the 27875 metalloproteinase gene associated with a dysfunction provides a diagnostic tool for an active disease or susceptibility to disease when the disease results from overexpression, underexpression, or altered expression of an 27875 metalloproteinase.

Mutations in the 27875 metalloproteinase gene can be detected at the nucleic acid level by a variety of techniques. Genomic DNA can be analyzed directly or can be amplified by using PCR prior to analysis. RNA or cDNA can be used in the same way.

In certain embodiments, detection of the mutation involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran *et al.* (1988) *Science* 241:1077-1080; and Nakazawa *et al.* (1994) *PNAS* 91:360-364), the latter of which can be particularly useful

for detecting point mutations in the gene (see Abravaya *et al.* (1995) *Nucleic Acids Res.* 23:675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically
5 hybridize to a gene under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. Deletions and insertions can be detected by a change in size of the amplified product compared to the normal genotype. Point mutations can be identified
10 by hybridizing amplified DNA to normal RNA or antisense DNA sequences.

It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication
15 (Guatelli *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi *et al.* (1988) *Bio/Technology* 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well-known to those of skill in the art. These detection schemes are
20 especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

Alternatively, mutations in an 27875 metalloproteinase gene can be directly identified, for example, by alterations in restriction enzyme digestion patterns determined by gel electrophoresis.

25 Further, sequence-specific ribozymes (U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

Perfectly matched sequences can be distinguished from mismatched sequences by nuclease cleavage digestion assays or by differences in melting temperature.

30 Sequence changes at specific locations can also be assessed by nuclease protection assays such as RNase and S1 protection or the chemical cleavage method.

Furthermore, sequence differences between a mutant 27875 metalloproteinase gene and a wild-type gene can be determined by direct DNA sequencing. A variety of

gradient gel electrophoresis (Myers *et al.* (1989) *Nature* 315:135). The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen *et al.* (1991) *Trends Genet.* 7:5). Examples of other techniques for detecting point mutations include, selective oligonucleotide hybridization, selective amplification, and selective primer extension.

In other embodiments, genetic mutations can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotide probes (Cronin *et al.* (1996) *Human Mutation* 7:244-255; Kozal *et al.* (1996) *Nature Medicine* 2:753-759). For example, genetic mutations can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin *et al. supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is

composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

27875 metalloproteinase polynucleotides are also useful for testing an individual for a genotype that while not necessarily causing the disease, nevertheless affects the treatment modality. Thus, the polynucleotides can be used to study the relationship between an individual's genotype and the individual's response to a compound used for treatment (pharmacogenomic relationship). In the present case, for example, a mutation in the 27875 metalloproteinase gene that results in altered affinity for zinc could result in an excessive or decreased drug effect with standard concentrations of zinc. Accordingly, the 27875 metalloproteinase polynucleotides described herein can be used to assess the mutation content of the gene in an individual in order to select an appropriate compound or dosage regimen for treatment.

Thus polynucleotides displaying genetic variations that affect treatment provide a diagnostic target that can be used to tailor treatment in an individual. Accordingly, the production of recombinant cells and animals containing these polymorphisms allow effective clinical design of treatment compounds and dosage regimens.

The methods can involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting mRNA, or genomic DNA, such that the presence of mRNA or genomic DNA is detected in the biological sample, and comparing the presence of mRNA or genomic DNA in the control sample with the presence of mRNA or genomic DNA in the test sample.

27875 metalloproteinase polynucleotides are also useful for chromosome identification when the sequence is identified with an individual chromosome and to a particular location on the chromosome. First, the DNA sequence is matched to the chromosome by *in situ* or other chromosome-specific hybridization. Sequences can also be correlated to specific chromosomes by preparing PCR primers that can be used for PCR screening of somatic cell hybrids containing individual chromosomes from the desired species. Only hybrids containing the chromosome containing the gene homologous to the primer will yield an amplified fragment. Sublocalization can be achieved using chromosomal fragments. Other strategies include prescreening with labeled flow-sorted chromosomes and preselection by hybridization to chromosome-specific libraries. Further mapping strategies include fluorescence *in situ* hybridization,

which allows hybridization with probes shorter than those traditionally used. Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on the chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

27875 metalloproteinase polynucleotides can also be used to identify individuals from small biological samples. This can be done for example using restriction fragment-length polymorphism (RFLP) to identify an individual. Thus, the polynucleotides described herein are useful as DNA markers for RFLP (See U.S. Patent No. 5,272,057).

Furthermore, the 27875 metalloproteinase sequence can be used to provide an alternative technique, which determines the actual DNA sequence of selected fragments in the genome of an individual. Thus, the 27875 metalloproteinase sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify DNA from an individual for subsequent sequencing.

Panels of corresponding DNA sequences from individuals prepared in this manner can provide unique individual identifications, as each individual will have a unique set of such DNA sequences. It is estimated that allelic variation in humans occurs with a frequency of about once per each 500 bases. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. 27875 metalloproteinase sequences can be used to obtain such identification sequences from individuals and from tissue. The sequences represent unique fragments of the human genome. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes.

If a panel of reagents from the sequences is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

27875 metalloproteinase polynucleotides can also be used in forensic identification procedures. PCR technology can be used to amplify DNA sequences taken from very small biological samples, such as a single hair follicle, body fluids (e.g. blood, saliva, or semen). The amplified sequence can then be compared to a standard
5 allowing identification of the origin of the sample.

27875 metalloproteinase polynucleotides can thus be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e. another DNA sequence that is
10 unique to a particular individual). As described above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to the noncoding region are particularly useful since greater polymorphism occurs in the noncoding regions, making it easier to differentiate individuals using this technique.

27875 metalloproteinase polynucleotides can further be used to provide polynucleotide reagents, e.g., labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue. This is useful in cases in which a forensic pathologist is presented with a tissue of unknown origin. Panels of 27875 metalloproteinase probes can be used to identify tissue by species
15 and/or by organ type.

In a similar fashion, these primers and probes can be used to screen tissue culture for contamination (i.e. screen for the presence of a mixture of different types of cells in a culture).

Alternatively, 27875 metalloproteinase polynucleotides can be used directly to
25 block transcription or translation of 27875 metalloproteinase gene sequences by means of antisense or ribozyme constructs. Thus, in a disorder characterized by abnormally high or undesirable 27875 metalloproteinase gene expression, nucleic acids can be directly used for treatment.

27875 metalloproteinase polynucleotides are thus useful as antisense constructs
30 to control 27875 metalloproteinase gene expression in cells, tissues, and organisms. A DNA antisense polynucleotide is designed to be complementary to a region of the gene involved in transcription, preventing transcription and hence production of 27875

include coding regions and particularly coding regions corresponding to the catalytic and other functional activities of the 27875 metalloproteinase protein.

15 27875 metalloproteinase polynucleotides also provide vectors for gene therapy in patients containing cells that are aberrant in 27875 metalloproteinase gene expression. Thus, recombinant cells, which include the patient's cells that have been engineered *ex vivo* and returned to the patient, are introduced into an individual where the cells produce the desired 27875 metalloproteinase protein to treat the individual.

20 The invention also encompasses kits for detecting the presence of an 27875 metalloproteinase nucleic acid in a biological sample. For example, the kit can comprise reagents such as a labeled or labelable nucleic acid or agent capable of detecting 27875 metalloproteinase nucleic acid in a biological sample; means for determining the amount of 27875 metalloproteinase nucleic acid in the sample; and means for comparing the
25 amount of 27875 metalloproteinase nucleic acid in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect 27875 metalloproteinase mRNA or DNA.

30 Computer Readable Means

 The nucleotide or amino acid sequences of the invention are also provided in a variety of mediums to facilitate use thereof. As used herein, "provided" refers to a manufacture, other than an isolated nucleic acid or amino acid molecule, which

contains a nucleotide or amino acid sequence of the present invention. Such a manufacture provides the nucleotide or amino acid sequences, or a subset thereof (e.g., a subset of open reading frames (ORFs)) in a form which allows a skilled artisan to examine the manufacture using means not directly applicable to examining the
5 nucleotide or amino acid sequences, or a subset thereof, as they exists in nature or in purified form.

In one application of this embodiment, a nucleotide or amino acid sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium that can be read and accessed
10 directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. The skilled artisan will readily appreciate how any of the presently known computer readable
15 mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide or amino acid sequence of the present invention.

As used herein, "recorded" refers to a process for storing information on computer readable medium. The skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate
20 manufactures comprising the nucleotide or amino acid sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide or amino acid sequence of the present invention. The choice of the data storage structure will
25 generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or
30 represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. The skilled artisan can readily adapt any number of dataprocessor structuring formats (e.g., text file or database) in order to obtain

computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

By providing the nucleotide or amino acid sequences of the invention in computer readable form, the skilled artisan can routinely access the sequence information for a variety of purposes. For example, one skilled in the art can use the nucleotide or amino acid sequences of the invention in computer readable form to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of the sequences of the invention which match a particular target sequence or target motif.

As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium for analysis and comparison to other sequences. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBIA).

For example, software which implements the BLAST (Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-410) and BLAZE (Brutlag *et al.* (1993) *Comp. Chem.* 17:203-207) search algorithms on a Sybase system can be used to identify open reading frames (ORFs) of the sequences of the invention which contain homology to ORFs or
5 proteins from other libraries. Such ORFs are protein encoding fragments and are useful in producing commercially important proteins such as enzymes used in various reactions and in the production of commercially useful metabolites.

Vectors/Host Cells

10 The invention also provides vectors containing 27875 metalloproteinase polynucleotides. The term "vector" refers to a vehicle, preferably a nucleic acid molecule that can transport 27875 metalloproteinase polynucleotides. When the vector is a nucleic acid molecule, the 27875 metalloproteinase polynucleotides are covalently linked to the vector nucleic acid. With this aspect of the invention, the vector includes a
15 plasmid, single or double stranded phage, a single or double stranded RNA or DNA viral vector, or artificial chromosome, such as a BAC, PAC, YAC, OR MAC.

A vector can be maintained in the host cell as an extrachromosomal element where it replicates and produces additional copies of 27875 metalloproteinase polynucleotides. Alternatively, the vector may integrate into the host cell genome and
20 produce additional copies of 27875 metalloproteinase polynucleotides when the host cell replicates.

The invention provides vectors for the maintenance (cloning vectors) or vectors for expression (expression vectors) of 27875 metalloproteinase polynucleotides. The vectors can function in procaryotic or eukaryotic cells or in both (shuttle vectors).

25 Expression vectors contain cis-acting regulatory regions that are operably linked in the vector to 27875 metalloproteinase polynucleotides such that transcription of the polynucleotides is allowed in a host cell. The polynucleotides can be introduced into the host cell with a separate polynucleotide capable of affecting transcription. Thus, the second polynucleotide may provide a trans-acting factor interacting with the cis-
30 regulatory control region to allow transcription of 27875 metalloproteinase polynucleotides from the vector. Alternatively, a trans-acting factor may be supplied by the host cell. Finally, a trans-acting factor can be produced from the vector itself.

It is understood, however, that in some embodiments, transcription and/or translation of 27875 metalloproteinase polynucleotides can occur in a cell-free system.

The regulatory sequence to which the polynucleotides described herein can be operably linked include promoters for directing mRNA transcription. These include, but
5 are not limited to, the left promoter from bacteriophage λ , the lac, TRP, and TAC promoters from *E. coli*, the early and late promoters from SV40, the CMV immediate early promoter, the adenovirus early and late promoters, and retrovirus long-terminal repeats.

In addition to control regions that promote transcription, expression vectors may
10 also include regions that modulate transcription, such as repressor binding sites and enhancers. Examples include the SV40 enhancer, the cytomegalovirus immediate early enhancer, polyoma enhancer, adenovirus enhancers, and retrovirus LTR enhancers.

In addition to containing sites for transcription initiation and control, expression vectors can also contain sequences necessary for transcription termination and, in the
15 transcribed region a ribosome binding site for translation. Other regulatory control elements for expression include initiation and termination codons as well as polyadenylation signals. The person of ordinary skill in the art would be aware of the numerous regulatory sequences that are useful in expression vectors. Such regulatory sequences are described, for example, in Sambrook *et al.* (1989) *Molecular Cloning: A*
20 *Laboratory Manual 2nd. ed.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

A variety of expression vectors can be used to express an 27875 metalloproteinase polynucleotide. Such vectors include chromosomal, episomal, and virus-derived vectors, for example vectors derived from bacterial plasmids, from
25 bacteriophage, from yeast episomes, from yeast chromosomal elements, including yeast artificial chromosomes, from viruses such as baculoviruses, papovaviruses such as SV40, Vaccinia viruses, adenoviruses, poxviruses, pseudorabies viruses, and retroviruses. Vectors may also be derived from combinations of these sources such as those derived from plasmid and bacteriophage genetic elements, e.g. cosmids and
30 phagemids. Appropriate cloning and expression vectors for prokaryotic and eukaryotic hosts are described in Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual 2nd. ed.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

The regulatory sequence may provide constitutive expression in one or more host cells (i.e. tissue specific) or may provide for inducible expression in one or more cell types such as by temperature, nutrient additive, or exogenous factor such as a hormone or other ligand. A variety of vectors providing for constitutive and inducible expression
5 in prokaryotic and eukaryotic hosts are well known to those of ordinary skill in the art.

27875 metalloproteinase polynucleotides can be inserted into the vector nucleic acid by well-known methodology. Generally, the DNA sequence that will ultimately be expressed is joined to an expression vector by cleaving the DNA sequence and the expression vector with one or more restriction enzymes and then ligating the fragments
10 together. Procedures for restriction enzyme digestion and ligation are well known to those of ordinary skill in the art.

The vector containing the appropriate polynucleotide can be introduced into an appropriate host cell for propagation or expression using well-known techniques. Bacterial cells include, but are not limited to, *E. coli*, *Streptomyces*, and *Salmonella*
15 *typhimurium*. Eukaryotic cells include, but are not limited to, yeast, insect cells such as *Drosophila*, animal cells such as COS and CHO cells, and plant cells.

As described herein, it may be desirable to express the polypeptide as a fusion protein. Accordingly, the invention provides fusion vectors that allow for the production of 27875 metalloproteinase polypeptides. Fusion vectors can increase the expression of
20 a recombinant protein, increase the solubility of the recombinant protein, and aid in the purification of the protein by acting for example as a ligand for affinity purification. A proteolytic cleavage site may be introduced at the junction of the fusion moiety so that the desired polypeptide can ultimately be separated from the fusion moiety. Proteolytic enzymes include, but are not limited to, factor Xa, thrombin, and enterokinase. Typical
25 fusion expression vectors include pGEX (Smith *et al.* (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-
30 fusion *E. coli* expression vectors include pTrc (Amann *et al.* (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.* (1990) *Gene Expression Technology: Methods in Enzymology* 185:60-89).

Recombinant protein expression can be maximized in a host bacteria by providing a genetic background wherein the host cell has an impaired capacity to

expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith *et al.* (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow *et al.* 15 (1989) *Virology* 170:31-39).

In certain embodiments of the invention, the polynucleotides described herein are expressed in mammalian cells using mammalian expression vectors. Examples of mammalian expression vectors include pCDM8 (Seed, B. (1987) *Nature* 329:840) and pMT2PC (Kaufman *et al.* (1987) *EMBO J.* 6:187-195).

20 The expression vectors listed herein are provided by way of example only of the well-known vectors available to those of ordinary skill in the art that would be useful to express 27875 metalloproteinase polynucleotides. The person of ordinary skill in the art would be aware of other vectors suitable for maintenance propagation or expression of the polynucleotides described herein. These are found for example in Sambrook *et al.* 25 (1989) *Molecular Cloning: A Laboratory Manual 2nd, ed., Cold Spring Harbor Laboratory*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

The invention also encompasses vectors in which the nucleic acid sequences described herein are cloned into the vector in reverse orientation, but operably linked to a regulatory sequence that permits transcription of antisense RNA. Thus, an antisense 30 transcript can be produced to all, or to a portion, of the polynucleotide sequences described herein, including both coding and non-coding regions. Expression of this antisense RNA is subject to each of the parameters described above in relation to

expression of the sense RNA (regulatory sequences, constitutive or inducible expression, tissue-specific expression).

The invention also relates to recombinant host cells containing the vectors described herein. Host cells therefore include prokaryotic cells, lower eukaryotic cells
5 such as yeast, other eukaryotic cells such as insect cells, and higher eukaryotic cells such as mammalian cells.

The recombinant host cells are prepared by introducing the vector constructs described herein into the cells by techniques readily available to the person of ordinary skill in the art. These include, but are not limited to, calcium phosphate transfection,
10 DEAE-dextran-mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, lipofection, and other techniques such as those found in Sambrook *et al.* (*Molecular Cloning: A Laboratory Manual*, 2d ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

15 Host cells can contain more than one vector. Thus, different nucleotide sequences can be introduced on different vectors of the same cell. Similarly, 27875 metalloproteinase polynucleotides can be introduced either alone or with other polynucleotides that are not related to 27875 metalloproteinase polynucleotides such as those providing trans-acting factors for expression vectors. When more than one vector
20 is introduced into a cell, the vectors can be introduced independently, co-introduced or joined to the 27875 metalloproteinase polynucleotide vector.

In the case of bacteriophage and viral vectors, these can be introduced into cells as packaged or encapsulated virus by standard procedures for infection and transduction. Viral vectors can be replication-competent or replication-defective. In the case in which
25 viral replication is defective, replication will occur in host cells providing functions that complement the defects.

Vectors generally include selectable markers that enable the selection of the subpopulation of cells that contain the recombinant vector constructs. The marker can be contained in the same vector that contains the polynucleotides described herein or
30 may be on a separate vector. Markers include tetracycline or ampicillin-resistance genes for prokaryotic host cells and dihydrofolate reductase or neomycin resistance for eukaryotic host cells. However, any marker that provides selection for a phenotypic trait will be effective.

15 chromatography, or high performance liquid chromatography.

It is also understood that depending upon the host cell in recombinant production of the polypeptides described herein, the polypeptides can have various glycosylation patterns, depending upon the cell, or maybe non-glycosylated as when produced in bacteria. In addition, the polypeptides may include an initial modified methionine in
20 some cases as a result of a host-mediated process.

Uses of Vectors and Host Cells

It is understood that "host cells" and "recombinant host cells" refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell.
25 Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

The host cells expressing the polypeptides described herein, and particularly recombinant host cells, have a variety of uses. First, the cells are useful for producing
30 27875 metalloproteinase proteins or polypeptides that can be further purified to produce desired amounts of 27875 metalloproteinase protein or fragments. Thus, host cells containing expression vectors are useful for polypeptide production.

Host cells are also useful for conducting cell-based assays involving 27875 metalloproteinase or 27875 metalloproteinase fragments. Thus, a recombinant host cell expressing a native 27875 metalloproteinase is useful to assay for compounds that stimulate or inhibit 27875 metalloproteinase function. This includes zinc or peptide
5 binding, gene expression at the level of transcription or translation, and interaction with other cellular components.

Host cells are also useful for identifying 27875 metalloproteinase mutants in which these functions are affected. If the mutants naturally occur and give rise to a pathology, host cells containing the mutations are useful to assay compounds that have a
10 desired effect on the mutant 27875 metalloproteinase (for example, stimulating or inhibiting function) which may not be indicated by their effect on the native 27875 metalloproteinase.

Recombinant host cells are also useful for expressing the chimeric polypeptides described herein to assess compounds that activate or suppress activation by means of a
15 heterologous domain, segment, site, and the like, as disclosed herein.

Further, mutant 27875 metalloproteinases can be designed in which one or more of the various functions is engineered to be increased or decreased and used to augment or replace 27875 metalloproteinase proteins in an individual. Thus, host cells can provide a therapeutic benefit by replacing an aberrant 27875 metalloproteinase or
20 providing an aberrant 27875 metalloproteinase that provides a therapeutic result. In one embodiment, the cells provide 27875 metalloproteinases that are abnormally active.

In another embodiment, the cells provide 27875 metalloproteinases that are abnormally inactive. These 27875 metalloproteinases can compete with endogenous 27875 metalloproteinases in the individual.

25 In another embodiment, cells expressing 27875 metalloproteinases that cannot be activated, are introduced into an individual in order to compete with endogenous 27875 metalloproteinases for zinc, glycan, or peptide. For example, in the case in which excessive zinc is part of a treatment modality, it may be necessary to effectively inactivate zinc at a specific point in treatment. Providing cells that compete for the
30 molecule, but which cannot be affected by 27875 metalloproteinase activation would be beneficial.

Homologously recombinant host cells can also be produced that allow the *in situ* alteration of endogenous metalloproteinase polynucleotide sequences in a host cell

genome. The host cell includes, but is not limited to, a stable cell line, cell *in vivo*, or cloned microorganism. This technology is more fully described in WO 93/09222, WO 91/12650, WO 91/06667, U.S. 5,272,071, and U.S. 5,641,670. Briefly, specific polynucleotide sequences corresponding to the metalloproteinase polynucleotides or sequences proximal or distal to a metalloproteinase gene are allowed to integrate into a host cell genome by homologous recombination where expression of the gene can be affected. In one embodiment, regulatory sequences are introduced that either increase or decrease expression of an endogenous sequence. Accordingly, a metalloproteinase protein can be produced in a cell not normally producing it. Alternatively, increased expression of metalloproteinase protein can be effected in a cell normally producing the protein at a specific level. Further, expression can be decreased or eliminated by introducing a specific regulatory sequence. The regulatory sequence can be heterologous to the metalloproteinase protein sequence or can be a homologous sequence with a desired mutation that affects expression. Alternatively, the entire gene can be deleted. The regulatory sequence can be specific to the host cell or capable of functioning in more than one cell type. Still further, specific mutations can be introduced into any desired region of the gene to produce mutant metalloproteinase proteins. Such mutations could be introduced, for example, into the specific functional regions such as the peptide substrate-binding site.

In one embodiment, the host cell can be a fertilized oocyte or embryonic stem cell that can be used to produce a transgenic animal containing the altered metalloproteinase gene. Alternatively, the host cell can be a stem cell or other early tissue precursor that gives rise to a specific subset of cells and can be used to produce transgenic tissues in an animal. See also Thomas *et al.*, *Cell* 51:503 (1987) for a description of homologous recombination vectors. The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced gene has homologously recombined with the endogenous metalloproteinase gene is selected (see e.g., Li, E. *et al.* (1992) *Cell* 69:915). The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras (see e.g., Bradley, A. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E.J. Robertson, ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can

be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, A. (1991) *Current Opinion in Biotechnology* 2:823-829 and in PCT
5 International Publication Nos. WO 90/11354; WO 91/01140; and WO 93/04169.

The genetically engineered host cells can be used to produce non-human transgenic animals. A transgenic animal is preferably a mammal, for example a rodent, such as a rat or mouse, in which one or more of the cells of the animal include a transgene. A transgene is exogenous DNA which is integrated into the genome of a cell
10 from which a transgenic animal develops and which remains in the genome of the mature animal in one or more cell types or tissues of the transgenic animal. These animals are useful for studying the function of an 27875 metalloproteinase protein and identifying and evaluating modulators of 27875 metalloproteinase protein activity.

Other examples of transgenic animals include non-human primates, sheep, dogs,
15 cows, goats, chickens, and amphibians.

In one embodiment, a host cell is a fertilized oocyte or an embryonic stem cell into which 27875 metalloproteinase polynucleotide sequences have been introduced.

A transgenic animal can be produced by introducing nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing
20 the oocyte to develop in a pseudopregnant female foster animal. Any of the 27875 metalloproteinase nucleotide sequences can be introduced as a transgene into the genome of a non-human animal, such as a mouse.

Any of the regulatory or other sequences useful in expression vectors can form part of the transgenic sequence. This includes intronic sequences and polyadenylation
25 signals, if not already included. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the 27875 metalloproteinase protein to particular cells.

Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art
30 and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A

transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of transgenic mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene can further be bred to
5 other transgenic animals carrying other transgenes. A transgenic animal also includes animals in which the entire animal or tissues in the animal have been produced using the homologously recombinant host cells described herein.

In another embodiment, transgenic non-human animals can be produced which contain selected systems, which allow for regulated expression of the transgene. One
10 example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, e.g., Lakso *et al.* (1992) *PNAS* 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of *S. cerevisiae* (O'Gorman *et al.* (1991) *Science* 251:1351-1355. If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing
15 transgenes encoding both the *Cre* recombinase and a selected protein is required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be
20 produced according to the methods described in Wilmut *et al.* (1997) *Nature* 385:810-813 and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G₀ phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species
25 from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyst and then transferred to a pseudopregnant female foster animal. The offspring born of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

Transgenic animals containing recombinant cells that express the polypeptides
30 described herein are useful to conduct the assays described herein in an *in vivo* context. Accordingly, the various physiological factors that are present *in vivo* and that could affect binding or activation, may not be evident from *in vitro* cell-free or cell-based assays. Accordingly, it is useful to provide non-human transgenic animals to assay *in*

vivo 27875 metalloproteinase function, including peptide interaction, the effect of specific mutant 27875 metalloproteinases on 27875 metalloproteinase function and peptide interaction, and the effect of chimeric 27875 metalloproteinases. It is also possible to assess the effect of null mutations, that is mutations that substantially or
5 completely eliminate one or more 27875 metalloproteinase functions.

In general, methods for producing transgenic animals include introducing a nucleic acid sequence according to the present invention, the nucleic acid sequence capable of expressing the protein in a transgenic animal, into a cell in culture or *in vivo*. When introduced *in vivo*, the nucleic acid is introduced into an intact organism
10 such that one or more cell types and, accordingly, one or more tissue types, express the nucleic acid encoding the protein. Alternatively, the nucleic acid can be introduced into virtually all cells in an organism by transfecting a cell in culture, such as an embryonic stem cell, as described herein for the production of transgenic animals, and this cell can be used to produce an entire transgenic organism. As
15 described, in a further embodiment, the host cell can be a fertilized oocyte. Such cells are then allowed to develop in a female foster animal to produce the transgenic organism.

Pharmaceutical Compositions

20 27875 metalloproteinase nucleic acid molecules, proteins, modulators of the protein, and antibodies (also referred to herein as "active compounds") can be incorporated into pharmaceutical compositions suitable for administration to a subject, e.g., a human. Such compositions typically comprise the nucleic acid molecule, protein, modulator, or antibody and a pharmaceutically acceptable carrier.

25 The term "administer" is used in its broadest sense and includes any method of introducing the compositions of the present invention into a subject. This includes producing polypeptides or polynucleotides *in vivo* by *in vivo* transcription or translation of polynucleotides that have been exogenously introduced into a subject. Thus, polypeptides or nucleic acids produced in the subject from the exogenous compositions
30 are encompassed in the term "administer."

As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with

pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, such media can be used in the compositions of the invention. Supplementary active compounds can also be

5 incorporated into the compositions. A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application

10 can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and

15 agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous

20 solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy

25 syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can

30 be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid,

thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example,
5 aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (e.g., an 27875 metalloproteinase protein or anti- 27875 metalloproteinase antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally,
10 dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient
15 from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For oral administration, the agent can be contained in enteric forms to survive the stomach or further coated or mixed to be released in a particular region of the GI tract by known methods. For the
20 purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included
25 as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a
30 sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser, which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For
5 transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active
10 compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

15 In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid.
20 Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled
25 in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. "Dosage unit form" as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound
30 calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound

and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for
5 example, intravenous injection, local administration (U.S. 5,328,470) or by stereotactic injection (see e.g., Chen *et al.* (1994) *PNAS* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be
10 produced intact from recombinant cells, e.g. retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

As defined herein, a therapeutically effective amount of protein or polypeptide (i.e., an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, preferably about 0.01 to 25 mg/kg body weight, more preferably about 0.1 to 20
15 mg/kg body weight, and even more preferably about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight.

The skilled artisan will appreciate that certain factors may influence the dosage required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the
20 subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody can include a single treatment or, preferably, can include a series of treatments. In a preferred example, a subject is treated with antibody, protein, or polypeptide in the range of between about 0.1 to 20 mg/kg body weight, one time per week for between about 1
25 to 10 weeks, preferably between 2 to 8 weeks, more preferably between about 3 to 7 weeks, and even more preferably for about 4, 5, or 6 weeks. It will also be appreciated that the effective dosage of antibody, protein, or polypeptide used for treatment may increase or decrease over the course of a particular treatment. Changes in dosage may result and become apparent from the results of diagnostic assays as
30 described herein.

The present invention encompasses agents which modulate expression or activity. An agent may, for example, be a small molecule. For example, such small molecules include, but are not limited to, peptides, peptidomimetics, amino acids,

amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (i.e., including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds.

It is understood that appropriate doses of small molecule agents depends upon a number of factors within the ken of the ordinarily skilled physician, veterinarian, or researcher. The dose(s) of the small molecule will vary, for example, depending upon the identity, size, and condition of the subject or sample being treated, further depending upon the route by which the composition is to be administered, if applicable, and the effect which the practitioner desires the small molecule to have upon the nucleic acid or polypeptide of the invention. Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of subject or sample weight (e.g., about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. Such appropriate doses may be determined using the assays described herein. When one or more of these small molecules is to be administered to an animal (e.g., a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

5 This invention may be embodied in many different forms and should not be construed as limited to the embodiments set forth herein; rather, these embodiments are provided so that this disclosure will fully convey the invention to those skilled in the art. Many modifications and other embodiments of the invention will come to mind in one skilled in the art to which this invention pertains having the benefit of the teachings presented in the foregoing description. Although specific terms are employed, they are used as in the art unless otherwise indicated.

THAT WHICH IS CLAIMED:

1. An isolated polypeptide comprising an amino acid sequence selected
5 from the group consisting of:
- (a) the amino acid sequence shown in SEQ ID NO:1;
 - (b) the amino acid sequence of an allelic variant of the amino acid
sequence shown in SEQ ID NO:1;
 - (c) the amino acid sequence of a sequence variant of the amino acid
10 sequence shown in SEQ ID NO:1, wherein the sequence variant is encoded by a nucleic
acid molecule hybridizing to the nucleic acid molecule shown in SEQ ID NO:2 under
stringent conditions;
 - (d) a fragment of the amino acid sequence shown in SEQ ID NO:1,
wherein the fragment comprises at least 26 contiguous amino acids;
 - 15 (e) a fragment of residues 1-1545 of the amino acid sequence shown
in SEQ ID NO:1, wherein said fragment comprises at least 10 contiguous amino acids.
 - (f) the amino acid sequence of the mature 27875 metalloproteinase
polypeptide, comprising about amino acid residues 31-1687 of SEQ ID NO:1;
 - (g) an amino acid sequence comprising residues 244-259 of SEQ ID
20 NO:1;
 - (h) an amino acid sequence comprising residues 385-394 of SEQ ID
NO:1;
 - (i) an amino acid sequence comprising residues 541-592 of SEQ ID
NO:1;
 - 25 (j) an amino acid sequence comprising residues 542-592 of SEQ ID
NO:1;
 - (k) an amino acid sequence comprising residues 825-868 of SEQ ID
NO:1;
 - (l) an amino acid sequence comprising residues 949-988 of SEQ ID
30 NO:1;
 - (m) an amino acid sequence comprising residues 1415-1463 of SEQ
ID NO:1;

(n) an amino acid sequence comprising residues 1466-1521 of SEQ ID NO:1; and

(o) the amino acid sequence of an epitope bearing region of a polypeptide of (a)-(n).

5

2. An isolated polypeptide having at least 70% sequence identity with the amino acid sequence of SEQ ID NO:1.

3. The polypeptide a claim 2, wherein said polypeptide has at least 80% sequence identity with the amino acid sequence of SEQ ID NO:1.

10

4. The polypeptide a claim 3, wherein said polypeptide has at least 90% sequence identity with the amino acid sequence of SEQ ID NO:1.

15

5. The polypeptide a claim 4, wherein said polypeptide has at least 95% sequence identity with the amino acid sequence of SEQ ID NO:1.

6. An isolated antibody that selectively binds to a polypeptide of claim 1.

20

7. An isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of:

(a) the nucleotide sequence shown in SEQ ID NO:2;

(b) a nucleotide sequence encoding the amino acid sequence shown in SEQ ID NO:1;

25

(c) a nucleotide sequence encoding a polypeptide of claim 1; and

(d) a nucleotide sequence complementary to a nucleotide sequence of (a)-(c).

8. An isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of:

30

(a) a nucleotide sequence encoding a variant of the amino acid sequence shown in SEQ ID NO:1, wherein said nucleotide sequence hybridizes to the nucleotide sequence shown in SEQ ID NO:2 under stringent conditions;

- (b) a nucleotide sequence encoding the polypeptide of claim 2; and
- (c) a nucleotide sequence complementary to either of the nucleotide sequences in (a) or (b).

5 9. An isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of:

- (a) a nucleotide sequence encoding a fragment of the amino acid sequence shown in SEQ ID NO:1, wherein the fragment comprises at least 26 contiguous amino acids;
- 10 (b) a nucleotide sequence comprising at least 75 consecutive nucleotides of the sequence shown in SEQ ID NO:2;
- (c) a nucleotide sequence comprising at least 33 consecutive nucleotides of residues 1-4800 of SEQ ID NO:2;
- (d) a nucleotide sequence encoding residues 31-1687 of the amino
- 15 acid shown in SEQ ID NO:1;
- (e) a nucleotide sequence encoding residues 244-259 of SEQ ID NO:1;
- (f) a nucleotide sequence encoding residues 385-394 of SEQ ID NO:1;
- 20 (g) a nucleotide sequence encoding residues 541-592 of SEQ ID NO:1;
- (h) a nucleotide sequence encoding residues 542-592 of SEQ ID NO:1;
- (i) a nucleotide sequence encoding residues 825-868 of SEQ ID
- 25 NO:1;
- (j) a nucleotide sequence encoding residues 949-988 of SEQ ID NO:1;
- (k) a nucleotide sequence encoding residues 1415-1463 of SEQ ID NO:1; and
- 30 (l) a nucleotide sequence complementary to a nucleotide sequences of (a)-(k).

10. A nucleic acid vector comprising the nucleic acid sequences in any of claims 7-9.

11. A host cell containing the vector of claim 10.

12. A method for producing an 27875 metalloproteinase polypeptide, comprising: introducing a nucleotide sequence encoding a polypeptide sequence of claim 1 into a host cell, and culturing the host cell under conditions in which said polypeptide is expressed.

13. A method for detecting the presence of a polypeptide of claim 1 in a sample, comprising:

(a) contacting said sample with an agent that specifically allows detection of the presence of the polypeptide in the sample; and

(b) detecting the presence of the polypeptide.

14. The method of claim 13, wherein said agent specifically binds to said polypeptide.

15. The method of claim 14, wherein said agent is an antibody.

16. The method of claim 13, wherein said agent is a peptide substrate.

17. A kit comprising reagents used for the method of claim 13, wherein the reagents comprise an agent that specifically binds to said polypeptide.

18. A method for detecting the presence of a nucleic acid sequence of claims 7-9 in a sample, comprising: contacting the sample with an oligonucleotide that hybridizes to said nucleic acid sequence under stringent conditions and determining whether the oligonucleotide binds the nucleic acid sequence in said sample.

19. A kit comprising reagents used for the method of claim 18, wherein the reagents comprise said oligonucleotide.

20. A method for identifying an agent that interacts with a polypeptide of claim 1, comprising:

- 5 (a) contacting said polypeptide with an agent that interacts with said polypeptide; and
(b) detecting the interaction.

21. A method for identifying an agent that interacts with a nucleic acid sequence of claim 7, comprising:

- 10 (a) contacting said nucleic acid with an agent that interacts with said nucleic acid; and
(b) detecting the interaction.

22. A method for modulating the level or activity of a polypeptide of claim 1, comprising: contacting said polypeptide with an agent under conditions that allow the agent to modulate the activity of said polypeptide.

23. A method for modulating the level or activity of a nucleic acid of claim 7, comprising: contacting said nucleic with an agent under conditions that allow the agent to modulate the level or activity of said nucleic acid.

24. The method of claim 22 wherein said modulation is in an osteoblast.

25. The method of claim 24 wherein said modulation is in a human subject.

26. The method of claim 23 wherein said modulation is in an osteoblast.

27. The method of claim 26 where in said modulation is in a human subject.

30 28. A method for preventing or treating bone disorders, comprising:
administering a polypeptide of claim 1 to a subject in need of such treatment.

29. A non-human transgenic animal containing any of the nucleic acid sequences in claim 7.
30. A non-human transgenic animal expressing any of the polypeptides in claim 1.
31. A pharmaceutical composition containing any of the polypeptides in claim 1.
32. A pharmaceutical composition containing any of the nucleic acid molecules in claim 7.
33. An agent identified by the method of claim 20.
34. An agent identified by the method of claim 21.

Input file 27875cons; Output File 2787tra
Sequence length 5353

AAAGGGAATAAGCTTGCGGCCGCCCGGTTCCCTGCC	M	P	G	G	P	S	P	R	S	P	A		11
ATG CCC GGC GGC CCC AGT CCC CGC AGC CCC GCG													33
P L L R P L L L L L C A L A P G A F G P													31
CCT TTG CTG CGC CCC CTC CTC CTG CTC CTC TGC GCT CTG GCT CCC GGC GCC CCC GGA CCC													93
A P G R A T E G R A A L D I V H P V R V													51
GCA CCA GGA CGT GCA ACC GAG GGC CGG GCG GCA CTG GAC ATC GTG CAC CCG GTT CGA GTC													153
D A G G S F L S Y E L W P R A L R K R D													71
GAC GCG GGG GGC TCC TTC CTG TCC TAC GAG CTG TGG CCC CGC GCA CTG CGC AAG CGG GAT													213
V S V R R D A P A F Y E L Q Y R G R E L													91
GTA TCT GTG CGC CGA GAC GCG CCC GCC TTC TAC GAG CTA CAA TAC CGC GGG CGC GAG CTG													273
R F N L T A N Q H L L A P G F V S E T R													111
CGC TTC AAC CTG ACC GCC AAT CAG CAC CTG CTG GCG CCC GGC TTT GTG AGC GAG ACG CGG													333
R R G G L G R A H I R A H T P A C H L L													131
CGG CGC GGC GGC CTG GGC CGC GCG CAC ATC CGG GCC CAC ACC CCG GCC TGC CAC CTG CTT													393
G E V Q D P E L E G G L A A I S A C D G													151
GGC GAG GTG CAG GAC CCT GAG CTC GAG GGT GGC CTG GCG GCC ATC AGC GCC TGC GAC GGC													453
L K G V F Q L S N E D Y F I E P L D S A													171
CTG AAA GGT GTG TTC CAA CTC TCC AAC GAG GAC TAC TTC ATT GAG CCC CTG GAC AGT GCC													513
P A R P G H A Q P H V V Y K R Q A P E R													191
CCG GCC CGG CCT GGC CAC GCC CAG CCC CAT GTG GTG TAC AAG CGT CAG GCC CCG GAG AGG													573
L A Q R G D S S A P S T C G V Q V Y P E													211
CTG GCA CAG CGG GGT GAT TCC AGT GCT CCA AGC ACC TGT GGT GTG CAA GTG TAC CCA GAG													633
L E S R R E R W E Q R Q Q W R R P R L R													231
CTG GAG TCT CGA CGG GAG CGT TGG GAG CAG CGG CAG CAG TGG CGG CGG CCA CGG CTG AGG													693
R L H Q R S V S K E K W V E T L V V A D													251
CGT CTA CAC CAG CGG TCG GTC AGC AAA GAG AAG TGG GTG GAG ACC CTG GTA GTA GCT GAT													753
A K M V E Y H G Q P Q V E S Y V L T I M													271
GCC AAA ATG GTG GAG TAC CAC GGA CAG CCG CAG GTT GAG AGC TAT GTG CTG ACC ATC ATG													813
N M V A G L F H D P S I G N P I H I T I													291
AAC ATG GTG GCT GGC CTG TTT CAT GAC CCC AGC ATT GGG AAC CCC ATC CAC ATC ACC ATT													873
V R L V L L E D E E E D L K I T H H A D													311
GTG CGC CTG GTC CTG CTG GAA GAT GAG GAG GAG GAC CTA AAG ATC ACG CAC CAT GCA GAC													933
N T L K S F C K W Q K S I N M K G D A H													331
AAC ACC CTG AAG AGC TTC TGC AAG TGG CAG AAA AGC ATC AAC ATG AAG GGG GAT GCC CAT													993
P L H H D T A I L L T R K D L C A A M N													351
CCC CTG CAC CAT GAC ACT GCC ATC CTG CTC ACC AGA AAG GAC CTG TGT GCA GCC ATG AAC													1053
R P C E T L G L S H V A G M C Q F H R S													371
CGG CCC TGT GAG ACC CTG GGA CTG TCC CAT GTG GCG GGC ATG TGC CAG CCG CAC CGC AGC													1113
C S I N E D T G L P L A F T V A H E L G													391
TGC AGC ATC AAC GAG GAC ACG GGC CTG CCG CTG GCC TTC ACT GTA GCC CAC GAG CTC GGG													1173
H S F G I Q H D G S G N D C E P V G K R													411
CAC AGT TTT GGC ATT CAG CAT GAC GGA AGC GGC AAT GAC TGT GAG CCC GTT GGG AAA CGA													1233

FIG. 1A.

P	F	I	M	S	P	Q	L	L	Y	D	A	A	P	L	T	W	S	R	C	431
CCT	TTC	ATC	ATG	TCT	CCA	CAG	CTC	CTG	TAC	GAC	GCC	GCT	CCC	CTC	ACC	TGG	TCC	CGC	TGC	1293
S	R	Q	Y	I	T	R	F	L	D	R	G	W	G	L	C	L	D	D	P	451
AGC	CGC	CAG	TAT	ATC	ACC	AGG	TTC	CTT	GAC	CGT	GGG	TGG	GGC	CTG	TGC	CTG	GAC	GAC	CCT	1353
P	A	K	D	I	I	D	F	P	S	V	P	P	G	V	L	Y	D	V	S	471
CCT	GCC	AAG	GAC	ATT	ATC	GAC	TTC	CCC	TGC	GTG	CCA	CCT	GGC	GTC	CTC	TAT	GAT	GTA	AGC	1413
H	Q	C	R	L	Q	Y	G	A	Y	S	A	F	C	E	D	M	D	N	V	491
CAC	CAG	TGC	CGC	CTC	CAG	TAC	GGG	GCC	TAC	TCT	GCC	TTC	TGC	GAG	GAC	ATG	GAT	AAT	GTC	1473
C	H	T	L	W	C	S	V	G	T	T	C	H	S	K	L	D	A	A	V	511
TGC	CAC	ACA	CTC	TGG	TGC	TCT	GTG	GGG	ACC	ACC	TGT	CAC	TCC	AAG	CTG	GAT	GCA	GCC	GTG	1533
D	G	T	R	C	G	E	N	K	W	C	L	S	G	E	C	V	P	V	G	531
GAC	GGC	ACC	CGG	TGT	GGG	GAG	AAT	AAG	TGG	TGT	CTC	AGT	GGG	GAG	TGC	GTA	CCC	GTG	GGC	1593
F	R	P	E	A	V	D	G	G	W	S	G	W	S	A	W	S	I	C	S	551
TTC	CGG	CCC	GAG	GCC	GTG	GAT	GGT	GGC	TGG	TCT	GGC	TGG	AGC	GCC	TGG	TCC	ATC	TGC	TCA	1653
R	S	C	G	M	G	V	Q	S	A	E	R	Q	C	T	Q	P	T	P	K	571
CGG	AGC	TGT	GGC	ATG	GGC	GTA	CAG	AGC	GCC	GAG	CGG	CAG	TGC	ACG	CAG	CCT	ACG	CCC	AAA	1713
Y	K	G	R	Y	C	V	G	E	R	K	R	F	R	L	C	N	L	Q	A	591
TAC	AAA	GGC	AGA	TAC	TGT	GTG	GGT	GAG	CGC	AAG	CGC	TTC	CGC	CTC	TGC	AAC	CTG	CAG	GCC	1773
C	P	A	G	R	P	S	F	R	H	V	Q	C	S	H	F	D	A	M	L	611
TGC	CCT	GCT	GGC	CGC	CCC	TCC	TTC	CGC	CAC	GTC	CAG	TGC	AGC	CAC	TTT	GAC	GCT	ATG	CTC	1833
Y	K	G	Q	L	H	T	W	V	P	V	V	N	D	V	N	P	C	E	L	631
TAC	AAG	GGC	CAG	CTG	CAC	ACA	TGG	GTG	CCC	GTG	GTC	AAT	GAC	GTG	AAC	CCC	TGC	GAG	CTG	1893
H	C	K	P	A	N	E	Y	F	A	E	K	L	R	D	A	V	V	D	G	651
CAC	TGC	CGG	CCC	GCG	AAT	GAG	TAC	TTT	GCC	GAG	AAG	CTG	CGG	GAC	GCC	GTG	GTC	GAT	GGC	1953
T	P	C	Y	Q	V	R	A	S	R	D	L	C	I	N	G	I	C	K	N	671
ACC	CCC	TGC	TAC	CAG	GTC	CGA	GCC	AGC	CGG	GAC	CTC	TGC	ATC	AAC	GGC	ATC	TGT	AAG	AAC	2013
V	G	C	D	F	E	I	D	S	G	A	M	E	D	R	C	G	V	C	H	691
GTG	GGC	TGT	GAC	TTC	GAG	ATT	GAC	TCC	GGT	GCT	ATG	GAG	GAC	CGC	TGT	GGT	GTG	TGC	CAC	2073
G	N	G	S	T	C	H	T	V	S	G	T	F	E	E	A	E	G	L	G	711
GGC	AAC	GGC	TCC	ACC	TGC	CAC	ACC	GTG	AGC	GGG	ACC	TTC	GAG	GAG	GCC	GAG	GGC	CTG	GGG	2133
Y	V	D	V	G	L	I	P	A	G	A	R	E	I	R	I	Q	E	V	A	731
TAT	GTG	GAT	GTG	GGG	CTG	ATC	CCA	GCC	GGC	GCA	CGC	GAG	ATC	CGC	ATC	CAA	GAG	GTT	GCC	2193
E	A	A	N	F	L	A	L	R	S	E	D	P	E	K	Y	F	L	N	G	751
GAG	GCT	GCC	AAC	TTC	CTG	GCA	CTG	CGG	AGT	GAG	GAC	CCG	GAG	AAG	TAC	TTC	CTC	AAT	GGT	2253
G	W	T	I	Q	W	N	G	D	Y	Q	V	A	G	T	T	F	T	Y	A	772
GGC	TGG	ACC	ATC	CAG	TGG	AAC	GGG	GAC	TAC	CAG	GTG	GCA	GGG	ACC	ACC	TTC	ACA	TAC	GCA	2313
R	R	G	N	W	E	N	L	T	S	P	G	P	T	K	E	P	V	W	I	791
CGC	AGG	GGC	AAC	TGG	GAG	AAC	CTC	ACG	TCC	CCG	GGT	CCC	ACC	AAG	GAG	CCT	GTC	TGG	ATC	2373
Q	L	L	F	Q	E	S	N	P	G	V	H	Y	E	Y	T	I	H	R	E	811
CAG	CTG	CTG	TTC	CAG	GAG	AGC	AAC	CCT	GGG	GTG	CAC	TAC	GAG	TAC	ACC	ATC	CAC	AGG	GAG	2433
A	G	G	H	D	E	V	P	P	P	V	F	S	W	H	Y	G	P	W	T	831
GCA	GGT	GGC	CAC	GAC	GAG	GTC	CCG	CCG	CCC	GTG	TTC	TCC	TGG	CAT	TAT	GGG	CCC	TGG	ACC	2493
K	C	T	V	T	C	G	R	G	V	Q	R	Q	N	V	Y	C	L	E	R	851

FIG. 1B.

AAG	TGC	ACA	GTC	ACC	TGC	GGC	AGA	GGT	GTG	CAG	AGG	CAG	AAT	GTG	TAC	TGC	TTG	GAG	CGG	2553
Q	A	G	P	V	D	E	E	H	C	D	P	L	G	R	P	D	D	Q	Q	871
CAG	GCA	GGG	CCC	GTG	GAC	GAG	GAG	CAC	TGT	GAC	CCC	CTG	GGC	CGG	CCT	GAT	GAC	CAA	CAG	2613
R	K	C	S	E	Q	P	C	P	A	R	W	W	A	G	E	W	Q	L	C	891
AGG	AAG	TGC	AGC	GAG	CAG	CCC	TGC	CCT	GCC	AGG	TGG	TGG	GCA	GGT	GAG	TGG	CAG	CTG	TGC	2673
S	S	S	C	G	P	G	G	L	S	R	R	A	V	L	C	I	R	S	V	911
TCC	AGC	TCC	TGC	GGG	CCT	GGG	GGC	CTC	TCC	CGC	CGG	GCC	GTG	CTC	TGC	ATC	CGC	AGC	GTG	2733
G	L	D	E	Q	S	A	L	E	P	P	A	C	E	H	L	P	R	P	P	931
GGG	CTG	GAT	GAG	CAG	AGC	GCC	CTG	GAG	CCA	CCC	GCC	TGT	GAA	CAC	CTT	CCC	CGG	CCC	CCT	2793
T	E	T	P	C	N	R	H	V	P	C	P	A	T	W	A	V	G	N	W	951
ACT	GAA	ACC	CCT	TGC	AAC	CGC	CAT	GTA	CCC	TGT	CCG	GCC	ACC	TGG	GCT	GTG	GGG	AAC	TGG	2853
S	Q	C	S	V	T	C	G	E	G	T	Q	R	R	N	V	L	C	T	N	971
TCT	CAG	TGC	TCA	GTG	ACA	TGT	GGG	GAG	GGC	ACT	CAG	CGC	CGA	AAT	GTC	CTC	TGC	ACC	AAT	2913
D	T	G	V	P	C	D	E	A	Q	Q	P	A	S	E	V	T	C	S	L	991
GAC	ACC	GGT	GTC	CCC	TGT	GAC	GAG	GCC	CAG	CAG	CCA	GCC	AGC	GAA	GTC	ACC	TGC	TCT	CTG	2973
P	L	C	R	W	P	L	G	T	L	G	P	E	G	S	G	S	G	S	S	1011
CCA	CTC	TGT	CGG	TGG	CCC	CTG	GGC	ACA	CTG	GGC	CCT	GAA	GGC	TCA	GGC	AGC	GGC	TCC	TCC	3033
S	H	E	L	F	N	E	A	D	F	I	P	H	H	L	A	P	R	P	S	1031
AGC	CAC	GAG	CTC	TTC	AAC	GAG	GCT	GAC	TTC	ATC	CCG	CAC	CAC	CTG	GCC	CCA	CGC	CCT	TCA	3093
P	A	S	S	P	K	P	G	T	M	G	N	A	I	E	E	E	A	P	E	1051
CCC	GCC	TCA	TCA	CCC	AAG	CCA	GGC	ACC	ATG	GGC	AAC	GCC	ATT	GAG	GAG	GAG	GCT	CCA	GAG	3153
L	D	L	P	G	P	V	F	V	D	D	F	Y	Y	D	Y	N	F	I	N	1071
CTG	GAC	CTG	CCG	GGG	CCC	GTG	TTT	GTG	GAC	GAC	TTC	TAC	TAC	GAC	TAC	AAT	TTC	ATC	AAT	3213
F	H	E	D	L	S	Y	G	P	S	E	E	P	D	L	D	L	A	G	T	1091
TTC	CAC	GAG	GAT	CTG	TCC	TAC	GGG	CCC	TCT	GAG	GAG	CCC	GAT	CTA	GAC	CTG	GCG	GGG	ACA	3273
G	D	R	T	P	P	P	H	S	R	P	A	A	P	S	T	G	S	P	V	1111
GGG	GAC	CGG	ACA	CCC	CCA	CCA	CAC	AGC	CGT	CCT	GCT	GCG	CCC	TCC	ACG	GGT	AGC	CCT	GTG	3333
P	A	T	E	P	P	A	A	K	E	E	G	V	L	G	P	W	S	P	S	1131
CCT	GCC	ACA	GAG	CCT	CCT	GCA	GCC	AAG	GAG	GAG	GGG	GTA	CTG	GGA	CCT	TGG	TCC	CCG	AGC	3393
P	W	P	S	Q	A	G	R	S	P	P	P	P	S	E	Q	T	P	G	N	1151
CCT	TGG	CCT	AGC	CAG	GCC	GGC	CGC	TCC	CCA	CCC	CCA	CCC	TCA	GAG	CAG	ACC	CCT	GGG	AAC	3453
P	L	I	N	F	L	P	E	E	D	T	P	I	G	A	P	D	L	G	L	1171
CCT	TTG	ATC	AAT	TTC	CTG	CCT	GAG	GAA	GAC	ACC	CCC	ATA	GGG	GCC	CCA	GAT	CTT	GGG	CTC	3513
P	S	L	S	W	P	R	V	S	T	D	G	L	Q	T	P	A	T	P	E	1191
CCC	AGC	CTG	TCC	TGG	CCC	AGG	GTT	TCC	ACT	GAT	GGC	CTG	CAG	ACA	CCT	GCC	ACC	CCT	GAG	3573
S	Q	N	D	F	P	V	G	K	D	S	Q	S	Q	L	P	P	P	W	R	1211
AGC	CAA	AAT	GAT	TTC	CCA	GTT	GGC	AAG	GAC	AGC	CAG	AGC	CAG	CTG	CCC	CCT	CCA	TGG	CGG	3633
D	R	T	N	E	V	F	K	D	D	E	E	P	K	G	R	G	A	P	H	1231
GAC	AGG	ACC	AAT	GAG	GTT	TTC	AAG	GAT	GAT	GAG	GAA	CCC	AAG	GGC	CGC	GGA	GCA	CCC	CAC	3693
L	P	P	R	P	S	S	T	L	P	P	L	S	P	V	G	S	T	H	S	1251
CTG	CCC	CCG	AGA	CCC	AGC	TCC	ACG	CTG	CCC	CCT	TTG	TCC	CCT	GTT	GGC	AGC	ACC	CAC	TCC	3753
S	P	S	P	D	V	A	E	L	W	T	G	G	T	V	A	W	E	P	A	1271
TCT	CCT	AGT	CCT	GAC	GTG	GCG	GAG	CTG	TGG	ACA	GGA	GGC	ACA	GTG	GCC	TGG	GAG	CCA	GCT	3813

FIG. 1C.

L E G G L G P V D S E L W P T V G V A S 1291
CTG GAG GGT GGC CTG GGG CCT GTG GAC AGT GAA CTG TGG CCC ACT GTT GGG GTG GCT TCT 3873
L L P P P I A P L P E M K V R D S S L E 1311
CTC CTT CCT CCT CCC ATA GCC CCT CTG CCA GAG ATG AAG GTC AGG GAC AGT TCC CTG GAG 3933
P G T P S F P A P G P G S W D L Q T V A 1331
CCG GGG ACT CCC TCC TTC CCA GCC CCA GGA CCA GGC TCA TGG GAC CTG CAG ACT GTG GCA 3993
V W G T F L P T T L T G L G H M P E P A 1351
GTG TGG GGG ACC TTC CTC CCC ACA ACC CTG ACT GGC CTC GGG CAC ATG CCT GAG CCT GCC 4053
L N P G P K G Q P E S L T P E V P L S S 1371
CTG AAC CCA GGA CCC AAG GGT CAG CCT GAG TCC CTC ACC CCT GAG GTG CCC CTG AGC TCT 4113
R L L S T P A W D S P A N S H R V P E T 1391
AGG CTG CTG TCC ACA CCA GCT TGG GAC AGC CCC GCC AAC AGC CAC AGA GTC CCT GAG ACC 4173
Q P L A P S L A E A G P P A D P L V V R 1411
CAG CCG CTG GCT CCC AGC CTG GCT GAA GCG GGG CCC CCC GCG GAC CCG TTG GTT GTC AGG 4233
N A S W Q A G N W S E C S T T C G L G A 1431
AAC GCC AGC TGG CAA GCG GGA AAC TGG AGC GAG TGC TCT ACC ACC TGT GGC CTG GGT GCG 4293
V W K P V R C S S G R D E D C A P A G R 1451
GTC TGG AGG CCG GTG CGC TGT AGC TCC GGC CGG GAT GAG GAC TGC GCC CCC GCT GGC CGG 4353
P Q P A R R C H L R P C A T W H S G N W 1471
CCC CAG CCT GCC CGC CGC TGC CAC CTG CGG CCC TGT GCC ACC TGG CAC TCA GGC AAC TGG 4413
S K C S R S C G G G S S V R D V Q C V D 1491
AGT AAG TGC TCC CGC AGC TGC GGC GGA GGT TCC TCA GTG CGG GAC GTG CAG TGT GTG GAC 4473
T R D L R P L R P F H C Q P G P A K P P 1511
ACA CGG GAC CTC CGG CCA CTG CGG CCC TTC CAT TGT CAG CCC GGG CCT GCC AAG CCG CCT 4533
A H R P C G A Q P C L S W Y T S S W R E 1531
GCG CAC CGG CCC TGC GGG GCC CAG CCC TGC CTC AGC TGG TAC ACA TCT TCC TGG AGG GAG 4593
C S E A C G G G E Q Q R L V T C P E P G 1551
TGC TCC GAG GCC TGT GGC GGT GGT GAG CAG CAG CGT CTA GTG ACC TGC CCG GAG CCA GGC 4653
L C E E A L R P N T T R P C N T H P C T 1571
CTC TGC GAG GAG GCG CTG AGA CCC AAC ACC ACC CGG CCC TGC AAC ACC CAC CCC TGC ACG 4713
Q W V V G P W G Q C S A P C G G G V Q R 1591
CAG TGG GTG GTG GGG CCC TGG GGC CAG TGC TCA GCC CCC TGT GGT GGT GGT GTC CAG CGG 4773
R L V K C V N T Q T G L P E E D S D Q C 1611
CGC CTG GTC AAG TGT GTC AAC ACC CAG ACA GGG CTG CCC GAG GAA GAC AGT GAC CAG TGT 4833
G H E A W P E S S R P C G T L D C E P V 1631
GGC CAC GAG GCC TGG CCT GAG AGC TCC CGG CCG TGT GGC ACC GAG GAT TGT GAG CCC GTC 4893
E P P R C E R D R L S F G F C E T L R L 1651
GAG CCT CCC CGC TGT GAG CGG GAC CGC CTG TCC TTC GGG TTC TGC GAG ACG CTG CGC CTA 4953
L G R C Q L P T I R T Q C C R S C S P P 1671
CTG GGC CGC TGC CAG CTG CCC ACC ATC CGC ACC CAG TGC TGC CGC TCG TGC TCT CCG CCC 5013
S H G A P S R G H Q R V A R R * 1687
AGC CAC GGC GCC CCC TCC CGA GGC CAT CAG CGG GTT GCC CGC CGC TGA 5061

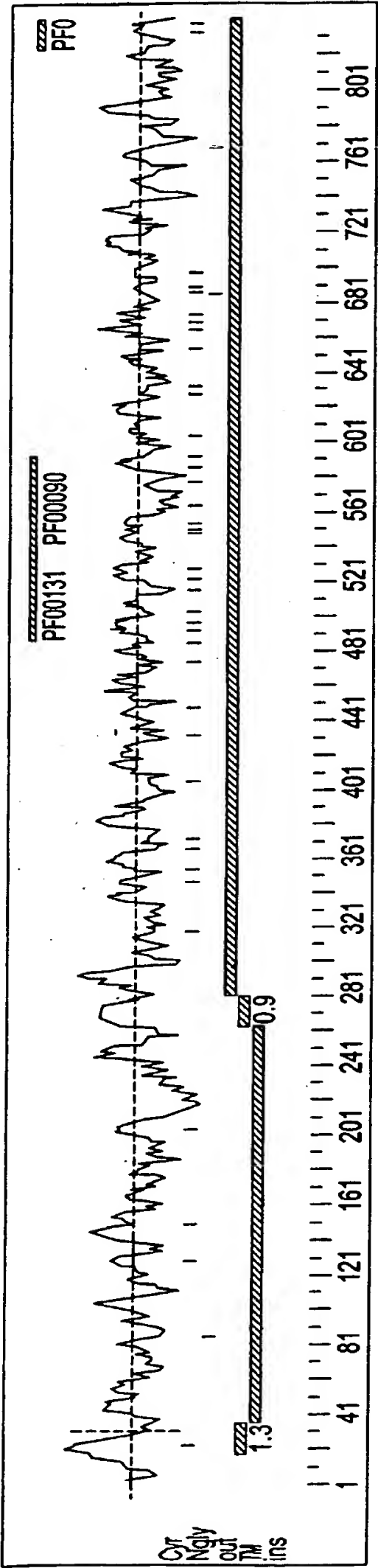
CTGTGCCAGGATGCACAGACCGACCGACAGACCTCAGTGGCCACCACGGGCTGTGGCGGAGCTCCCGCCCCCTGCGCCC

FIG. 1D.

TAATGGTGCTAACCCCTCTCACTACCCAGCAGCAGGCTGGGGACCTCCTCCCCCTCAAAAAGGTATTTTTTTATTCT
AACAGTTTGTGTAACATTTATTAGATTTTACATAAATGAGCATCTACCAAAAAAAAAAAAAAGGGCGGCCGCTAGAC
TAGTCTAGAGAAAAACCTC

FIG. 1E.

Analysis of 27875 (1687 aa)

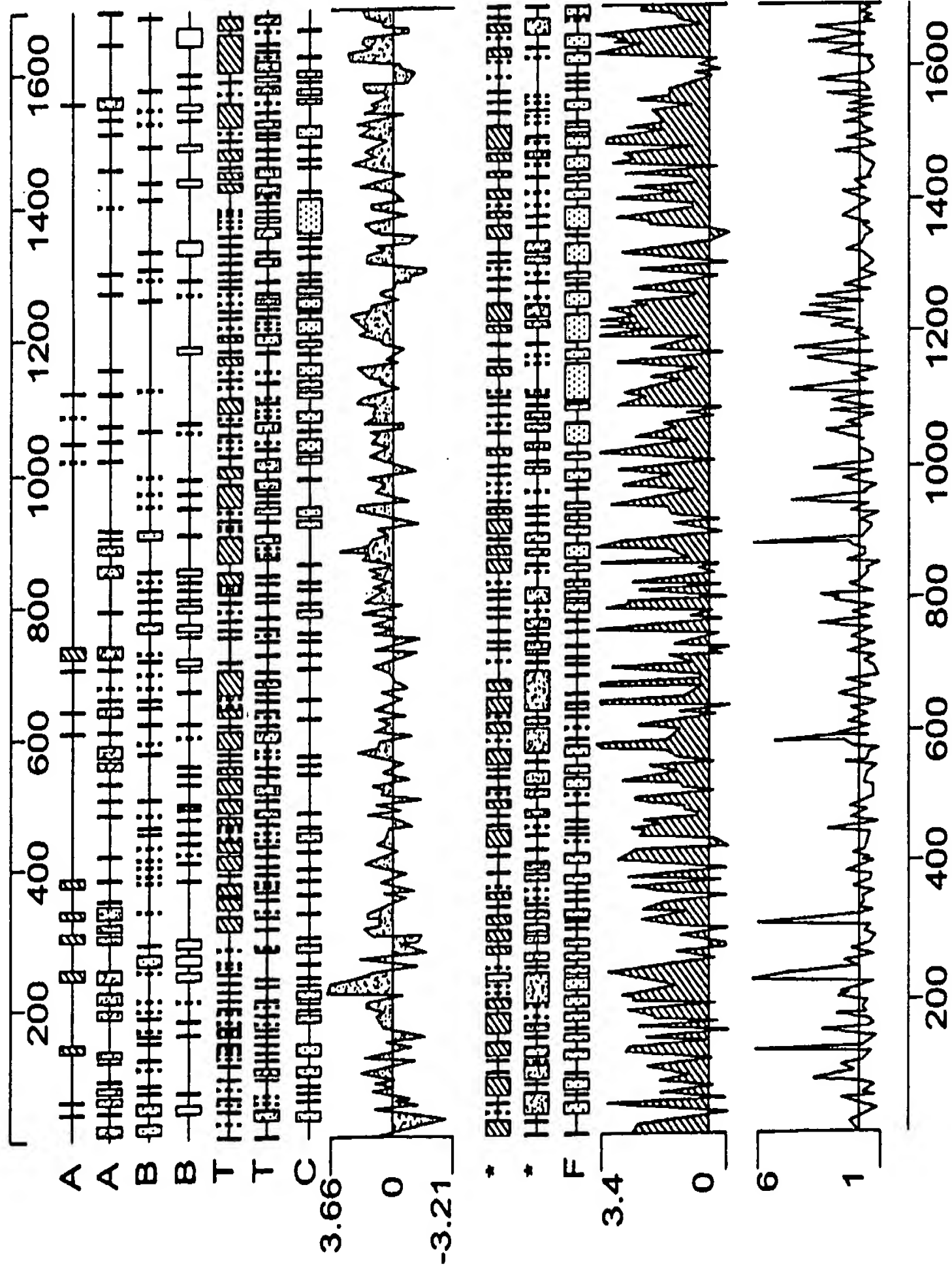


reprolysin: 242 - 452
metallothrombospondin 488 - 567

FIG. 2.

- ☒ Alpha, Regions
- ☒ Alpha, Regions
- ☒ Beta, Regions
- ☐ Beta, Regions
- ☒ Turn, Regions
- ☒ Turn, Regions
- ☒ Coil, Regions
- ☒ Hydrophilicity
- ☒ Alpha, Amphip
- ☒ Beta, Amphip
- ☒ Flexible Region
- ☒ Antigenic Inde
- ☐ Surface Proba

FIC



SUBSTITUTE SHEET (RULE 26)

Signal Peptide Predictions for 27875

Method	Predict	Score	Mat@
SignalP (eukaryote)	YES		30

Note: amino-terminal 70aa used for signal peptide prediction

Transmembrane Segments Predicted by MEMSAT

Start	End	Orient	Score
1573	1589	out->ins	0.6

Transmembrane segments for presumed mature peptide

Start	End	Orient	Score
237	254	ins->out	0.3

Prosite Pattern Matches for 27875

Prosite version: Release 12.2 of February 1995

-PS00001/PDOC00001/ASN_GLYCOSYLSTION N-glycosylation site.

Query: 94 NLTA 97
Query: 693 NGST 696
Query: 776 NLTS 781

FIG. 4A.

Query: 971 NDTG 974
 Query: 1412 NASW 1415
 Query: 1419 NWSE 1415
 Query: 1470 NWSK 1415
 Query: 1560 NDTG 1415

PS00002/PDDC00002/GLYCOSAMINOGLYCAN Glycosaminoglycan attachment site.

RU Additional rules:

RU There must be at least two acidic amino acids (Glu or Asp) from -2 to
 RU -4 relative to the serine.

Query: 1006 SGSG 1009

PS00004/PDDC00004/CAMP_PHOSPHO_SITE cAMP- and cGMP-dependent protein kinase phosphorylation site.

Query: 872 RKCS 875

PS00005/PDDC00005/PKC_PHOSPHO_SITE Protein kinase C phosphorylation site.

Query: 6 SPR 8
 Query: 73 SVR 75
 Query: 110 TRR 112
 Query: 214 SRR 216
 Query: 313 TLK 315
 Query: 342 TRK 344
 Query: 569 TPK 571
 Query: 598 SFR 600
 Query: 901 SRR 903
 Query: 962 TQR 964
 Query: 1035 SPK 1037
 Query: 1370 SSR 1372
 Query: 1385 SHR 1387
 Query: 1440 SGR 1442
 Query: 1483 SVR 1485
 Query: 1528 SWR 1530
 Query: 1561 TTR 1563
 Query: 1619 SSR 1621
 Query: 1648 TLR 1650
 Query: 1659 TIR 1661

PS00006/PDDC00006/CK2_PHOSPHO_SITE Casein kinase II phosphorylation site.

Query: 147 SACD 150
 Query: 159 SNED 162
 Query: 214 SRRE 217
 Query: 342 TRKD 345
 Query: 371 SINE 376

FIG. 4B.

Query: 401	SGND	404
Query: 505	SKLD	508
Query: 605	SHFD	608
Query: 703	TFEE	706
Query: 917	SALE	920
Query: 957	TCGE	960
Query: 1011	SSHE	1014
Query: 1192	SQND	1195
Query: 1308	SSLE	1311
Query: 1397	SLAE	1400
Query: 1440	SGRD	1443
Query: 1483	SVRD	1486
Query: 1528	SWRE	1531
Query: 1546	TCPE	1549

-PS00007/PDDC00007/TYR_PHOSPHO_SITE Tyrosine kinase phosphorylation site.

Query: 740 RSEDPEKY 747

-PS00008/PDDC00008/MYRISTYL N-myristoylation site.

Query: 55	GSFLSY	60
Query: 115	GLGRAH	120
Query: 141	GGLAAI	146
Query: 379	GLPLAF	384
Query: 479	GAYSAF	484
Query: 513	GTRCGE	518
Query: 539	GGMSGW	544
Query: 557	GVQSAE	562
Query: 614	GQLHTW	619
Query: 667	GICKNV	672
Query: 688	GVCHGN	693
Query: 716	GILPAG	721
Query: 765	GTTFTY	770
Query: 774	GNWENL	779
Query: 1005	GSGSGS	1010
Query: 1039	GTMGNA	1044
Query: 1247	GSTHSS	1252
Query: 1263	GGTVAV	1268
Query: 1358	GQPESL	1363
Query: 1517	GAQPCL	1522
Query: 1624	GTEDCE	1629

PS00009/PDDC00009/AMIDATION Amidation site.

FIG. 4C.

Query: 408 VGKR 411

PS00016/PDOC00016/RGD Cell attachment sequence.

Query: 195 RGD 197

PS00142/PDOC00129/ZINC_PROTEASE Neutral zinc metallopeptidases, zinc-binding region signature.

Query: 385 TVAHELGHSF 394

PS00190/PDOC00169/CYTOCHROME_C Cytochrome c family heme-binding site signature.

Query: 687 CGVCHG 692

PS0025/PDOC00197/CRYSTALLIN_BETAGAMMA Crystalline beta and gamma 'Greek key' motif signature.

Query: 78 APAFYELQYRGRELRF 93

PS00340/PDOC00214/RECEPTOR_CYTOKINES_2 Growth factor and cytokines receptors family signature 2.

Query: 539 GGWSGWS 545

FIG. 4D.

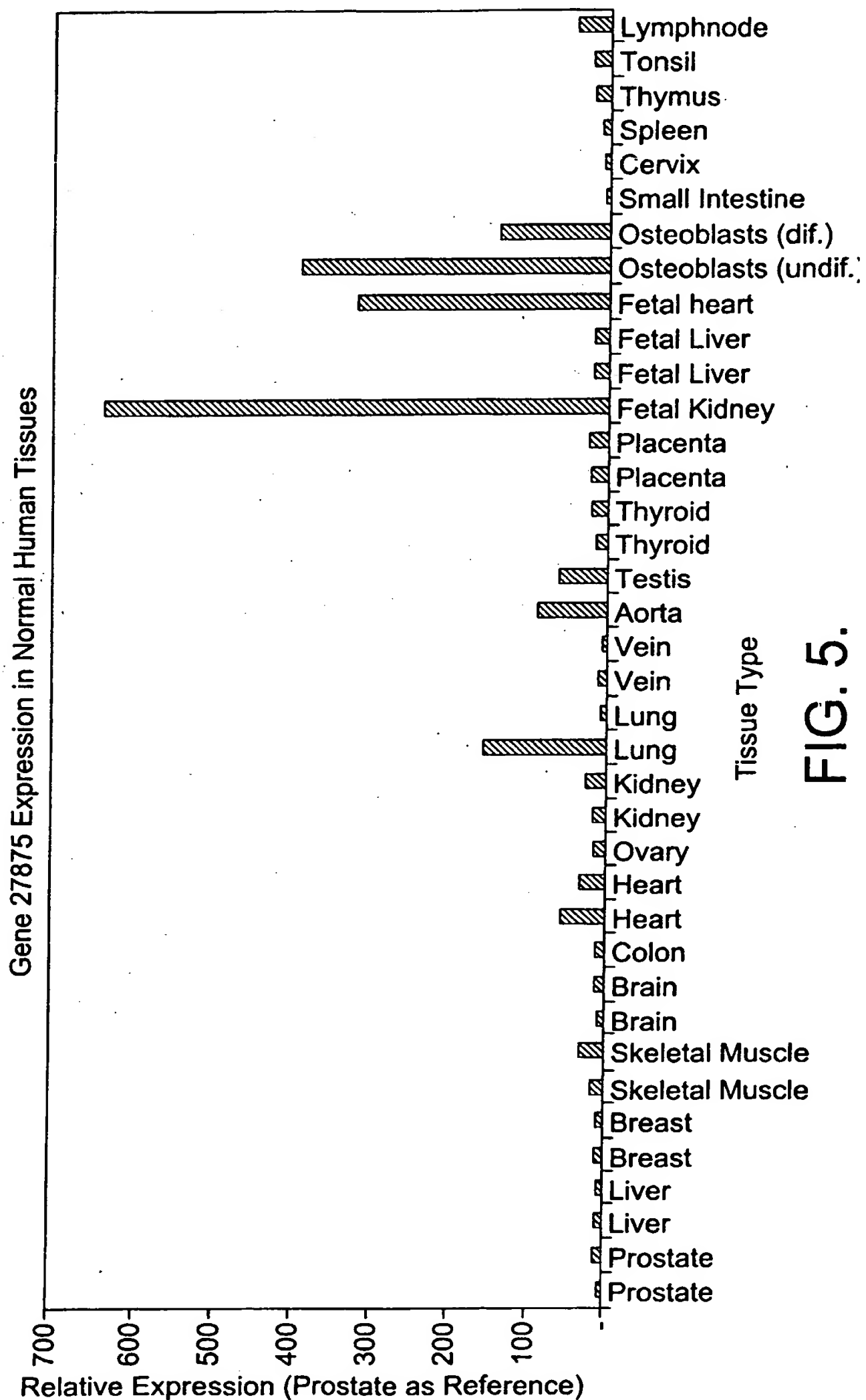
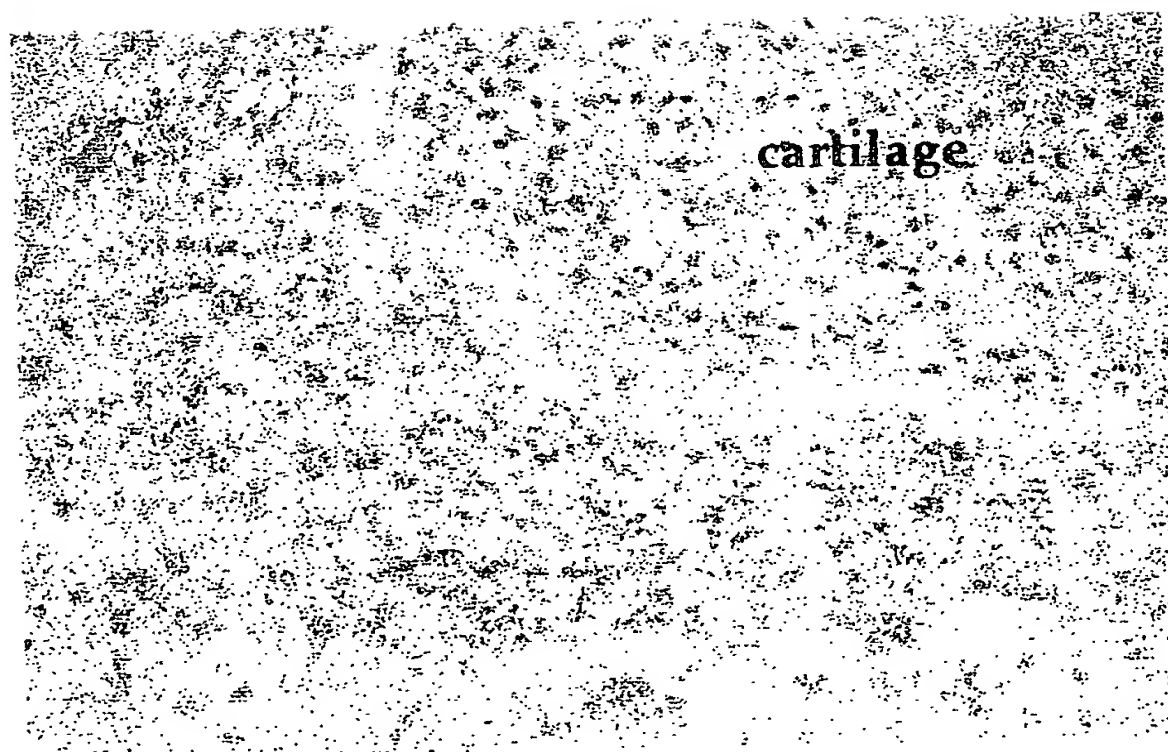
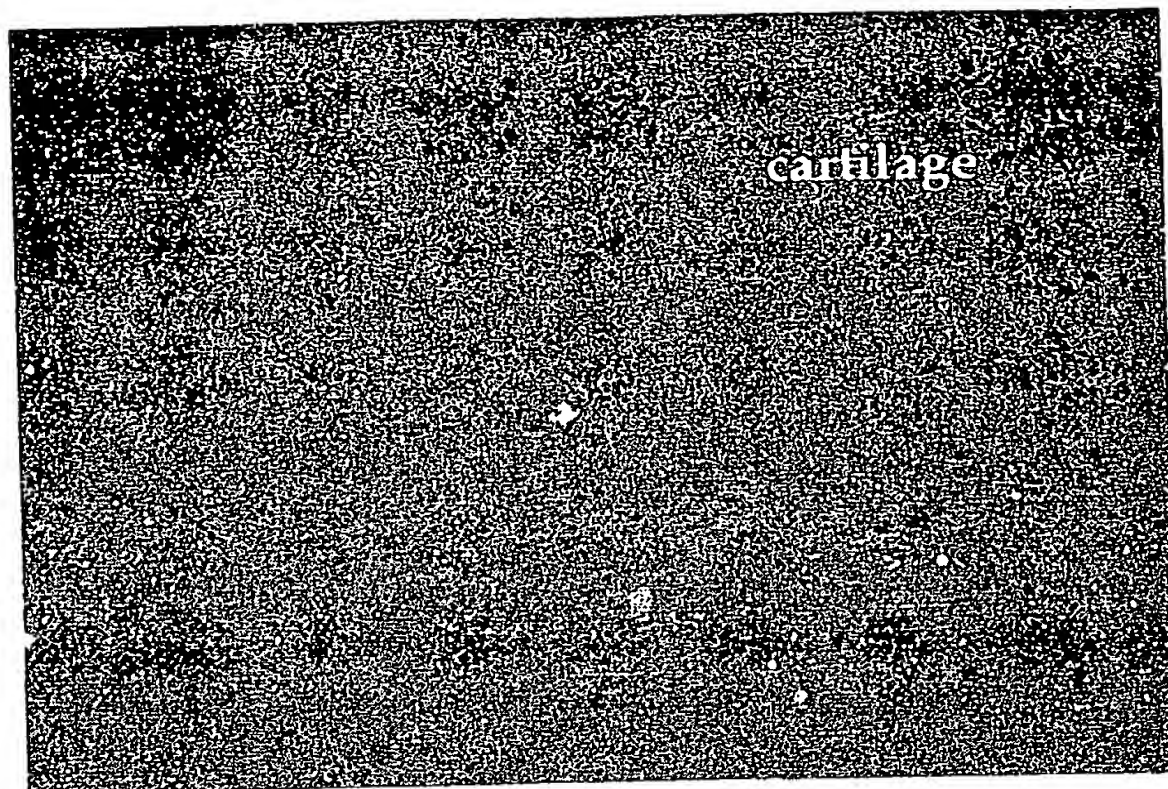


FIG. 5.

**Mining entry:27875
ADAM-TS**



20x

FIG. 6.

SUBSTITUTE SHEET (RULE 26)

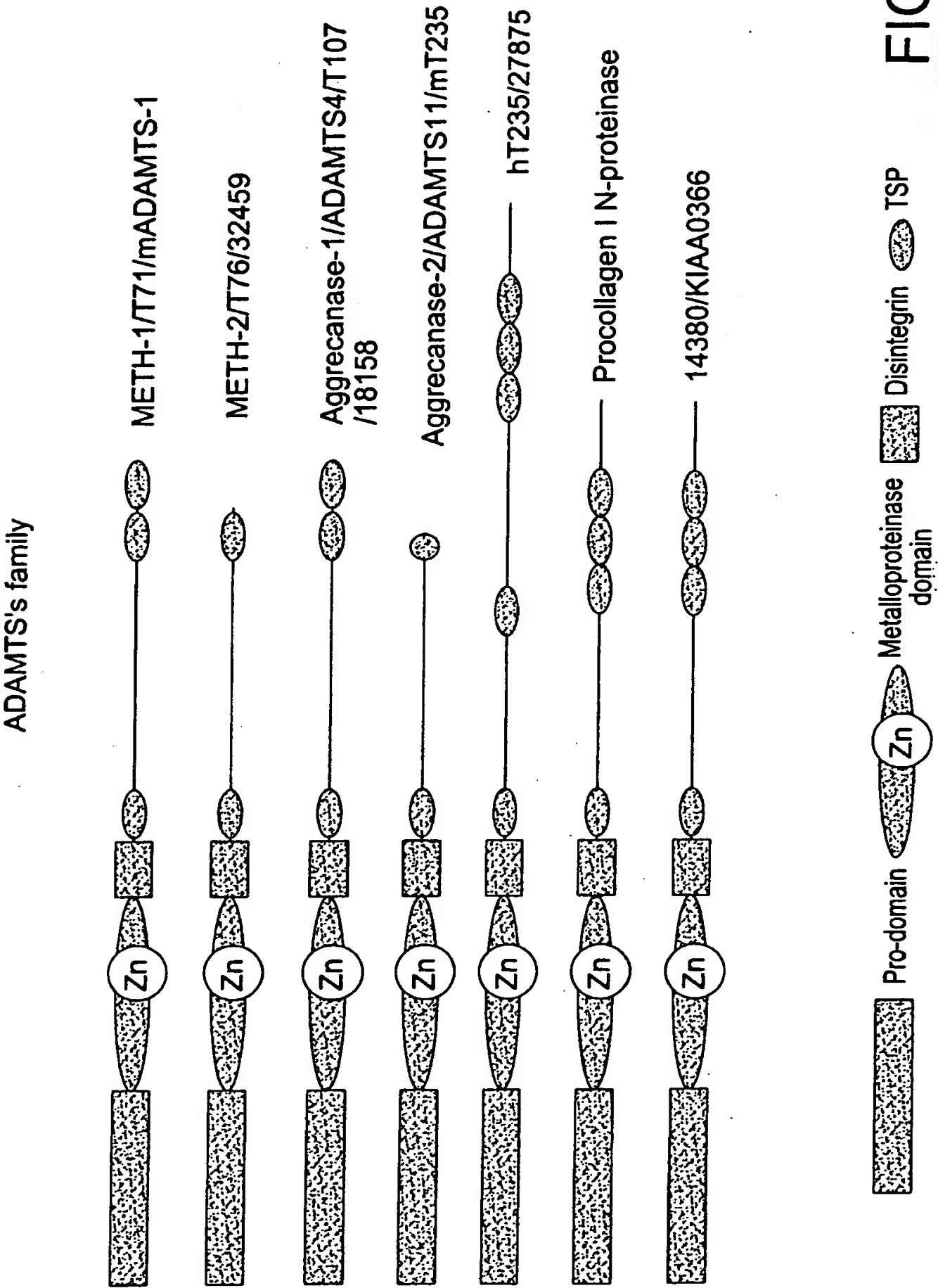


FIG. 7.

>42812
CGCGTCCGGG ACTCCCTCTG GCTTTACAAT TGCCCATGAG CTAGGACACA
GCTTCGGCAT CCAGCATGAT GGGAAAGAAA ATGACTGTGA GCCTGTGGGC
AGACATCCGT ACATCATGTC CCGCCAGCTC CAGTACGATC CCACTCCGCT
GACATGGTCC AAGTGCAGCG AGGAGTACAT CACCCGCTTC TTGGACCGAG
GCTGGGGGTT CTGTCTTGAT GACATACCTA AAAAGAAAGG CTTGAAGTCC
AAGGTCATTG CCCCCGGAGT GATCTATGAT GTTCACCACC AGTGCCAGCT
ACAATATGGA CCCAATGCTA CCTTCTGCAG GGAAGTAGAA AACGTCTGCC
AGACACTTGT GGTGGCTCCG TGAAGGGCTT TTGTCGCTCT AAGCTGGACG
CTGCTGCAGA TGGAA

>42812pep
ASGTPSGFTI AHELGHSGFI QHDGKENDCE PVGRHPYIMS RQLQYDPTPL
TWSKCSEFYI TRFLDRGWGF CLDDIPKKKG LSKSVIAPGV IYDVHHQCQL
QYGP NATFCR EVENVCQTLV VAP

FIG. 8.

	A	S	G	T	P	S		R	S	P	A
	M	P	G	G	P	S					
AAAGGGAATAAGCTTGCGGCCGCCCGGTTCCTGCC	ATG	CCC	GGC	GGC	CCC	AGT	CCC	CGC	AGC	CCC	GCG
P L L R P L L L L L	L C A L A P G A F G P	CCT TTG CTG CGC CCC CTC CTC CTG CTC CTC TGC GCT CTG GCT CCC GGC GCC CCC GGA CCC									
A P G R A T E G R A A L D I V H P V R V	GCA CCA GGA CGT GCA ACC GAG GGC CGG GCG GCA CTG GAC ATC GTG CAC CCG GTT CGA GTC										
D A G G S F L S Y E L W P R A L R K R D	GAC GCG GGG GGC TCC TTC CTG TCC TAC GAG CTG TGG CCC CGC GCA CTG CGC AAG CGG GAT										
V S V R R D A P A F Y E L Q Y R G R E L	GTA TCT GTG CGC CGA GAC GCG CCC GCC TTC TAC GAG CTA CAA TAC CGC GGG CGC GAG CTG										
R F N L T A N Q H L L A P G F V S E T R	CGC TTC AAC CTG ACC GCC AAT CAG CAC CTG CTG GCG CCC GGC TTT GTG AGC GAG ACG CGG										
R R G G L G R A H I R A H T P A C H L L	CGG CGC GGC GGC CTG GGC CGC GCG CAC ATC CGG GCC CAC ACC CCG GCC TGC CAC CTG CTT										
G E V Q D P E L E G G L A A I S A C D G	GGC GAG GTG CAG GAC CCT GAG CTC GAG GGT GGC CTG GCG GCC ATC AGC GCC TGC GAC GGC										
L K G V F Q L S N E D Y F I E P L D S A	CTG AAA GGT GTG TTC CAA CTC TCC AAC GAG GAC TAC TTC ATT GAG CCC CTG GAC AGT GCC										
P A R P G H A Q P H V V Y K R Q A P E R	CCG GCC CGG CCT GGC CAC GCC CAG CCC CAT GTG GTG TAC AAG CGT CAG GCC CCG GAG AGG										
L A Q R G D S S A P S T C G V Q V Y P E	CTG GCA CAG CGG GGT GAT TCC AGT GCT CCA AGC ACC TGT GGA GTG CAA GTG TAC CCA GAG										
L E S R R E R W E Q R Q Q W R R P R L R	CTG GAG TCT CGA CGG GAG CGT TGG GAG CAG CGG CAG CAG TGG CGG CGG CCA CGG CTG AGG										
R L H Q R S V S K E K W V E T L V V A D	CGT CTA CAC CAG CGG TCG GTC AGC AAA GAG AAG TGG GTG GAG ACC CTG GTA GTA GCT GAT										
A K M V E Y H G Q P Q V E S Y V L T I M	GCC AAA ATG GTG GAG TAC CAC GGA CAG CCG CAG GTT GAG AGC TAT GTG CTG ACC ATC ATG										
N M V A G L F H D P S I G N P I H I T I	AAC ATG GTG GCT GGC CTG TTT CAT GAC CCC AGC ATT GGG AAC CCC ATC CAC ATC ACC ATT										
V R L V L L E D E E E D L K I T H H A D	GTG CGC CTG GTC CTG CTG GAA GAT GAG GAG GAG GAC CTA AAG ATC ACG CAC CAT GCA GAC										
N T L K S F C K W Q K S I N M K G D A H	AAC ACC CTG AAG AGC TTC TGC AAG TGG CAG AAA AGC ATC AAC ATG AAG GGG GAT GCC CAT										
P L H H D T A I L L T R K D L C A A M N	CCC CTG CAC CAT GAC ACT GCC ATC CTG CTC ACC AGA AAG GAC CTG TGT GCA GCC ATG AAC										
R P C E T L G L S H V A G M C Q P H R S	CGG CCC TGT GAG ACC CTG GGA CTG TCC CAT GTG GCG GGC ATG TGC CAG CCG CAC CGC AGC										
			G	F	T	I	A	H	E	L	G
C S I N E D T G L P L A F T V A H E L G	TGC AGC ATC AAC GAG GAC ACG GGC CTG CCG CTG GCC TTC ACT GTA GCC CAC GAG CTC GGG										
H S F G I Q H D G K E N D C E P V G R H	K S F G I Q H D G S G N D C E P V G K R										
CAC AGT TTT GGC ATT CAG CAT GAC GGA AGC GGC AAT GAC TGT GAG CCC GTT GGG AAA CGA											

FIG. 9A.

P	Y	I	M	S	R	Q	L	Q	Y	D	P	T	P	L	T	W	S	K	C		431
P	F	I	M	S	P	Q	L	L	Y	D	A	A	P	L	T	W	S	R	C		1293
CCT	TTC	ATC	ATG	TCT	CCA	CAG	CTC	CTG	TAC	GAC	GCC	GCT	CCC	CTC	ACC	TGG	TCC	CGC	TGC		
S	E	E	Y	I	T	R	F	L	D	R	G	W	G	F	C	L	D	D	I		451
S	R	Q	Y	I	T	R	F	L	D	R	G	W	G	L	C	L	D	D	P		1353
AGC	CGC	CAG	TAT	ATC	ACC	AGG	TTC	CTT	GAC	CGT	GGG	TGG	GGC	CTG	TGC	CTG	GAC	GAC	CCT		
P	K	K	K	G	L	K	S	K	V	I	A	P	G	V	I	Y	D	V	H		471
P	A	K	D	I	I	D	F	P	S	V	P	P	G	V	L	Y	D	V	S		1413
CCT	GCC	AAG	GAC	ATT	ATC	GAC	TTC	CCC	TCG	GTG	CCA	CCT	GGC	GTC	CTC	TAT	GAT	GTA	AGC		
H	Q	C	Q	L	Q	Y	G	P	N	A	T	F	C	R	E	V	E	N	V		491
H	Q	C	R	L	Q	Y	G	A	Y	S	A	F	C	E	D	M	D	N	V		1473
CAC	CAG	TGC	CGC	CTC	CAG	TAC	GGG	GCC	TAC	TCT	GCC	TTC	TGC	GAG	GAC	ATG	GAT	AAT	GTC		
C	Q	T	L																		511
C	H	T	L	W	C	S	V	G	T	T	C	H	S	K	L	D	A	A	V		1533
TGC	CAC	ACA	CTC	TGG	TGC	TCT	GTG	GGG	ACC	ACC	TGT	CAC	TCC	AAG	CTG	GAT	GCA	GCC	GTG		531
D	G	T	R	C	G	E	N	K	W	C	L	S	G	E	C	V	P	V	G		1593
GAC	GGC	ACC	CGG	TGT	GGG	GAG	AAT	AAG	TGG	TGT	CTC	AGT	GGG	GAG	TGC	GTA	CCC	GTG	GGC		551
F	R	P	E	A	V	D	G	G	W	S	G	W	S	A	W	S	I	C	S		1653
TTC	CGG	CCC	GAG	GCC	GTG	GAT	GGT	GGC	TGG	TCT	GGC	TGG	AGC	GCC	TGG	TCC	ATC	TGC	TCA		571
R	S	C	G	M	G	V	Q	S	A	E	R	Q	C	T	Q	P	T	F	K		1713
CGG	AGC	TGT	GGC	ATG	GGC	GTA	CAG	AGC	GCC	GAG	CGG	CAG	TGC	ACG	CAG	CCT	ACG	CCC	AAA		591
Y	K	G	R	Y	C	V	G	E	R	K	R	F	R	L	C	N	L	Q	A		1773
TAC	AAA	GGC	AGA	TAC	TGT	GTG	GGT	GAG	CGC	AAG	CGC	TTC	CGC	CTC	TGC	AAC	CTG	CAG	GCC		611
C	P	A	G	R	P	S	F	R	H	V	Q	C	S	H	F	D	A	M	L		1833
TGC	CCT	GCT	GGC	CGC	CCC	TCC	TTC	CGC	CAC	GTC	CAG	TGC	AGC	CAC	TTT	GAC	GCT	ATG	CTC		631
Y	K	G	Q	L	H	T	W	V	P	V	V	N	D	V	N	P	C	E	L		1893
TAC	AAG	GGC	CAG	CTG	CAC	ACA	TGG	GTG	CCC	GTG	GTC	AAT	GAC	GTG	AAC	CCC	TGC	GAG	CTG		651
H	C	K	P	A	N	E	Y	F	A	E	K	L	R	D	A	V	V	D	G		1953
CAC	TGC	CGG	CCC	GCG	AAT	GAG	TAC	TTT	GCC	GAG	AAG	CTG	CGG	GAC	GCC	GTG	GTC	GAT	GGC		671
T	P	C	Y	Q	V	R	A	S	R	D	L	C	I	N	G	I	C	K	N		2013
ACC	CCC	TGC	TAC	CAG	GTC	CGA	GCC	AGC	CGG	GAC	CTC	TGC	ATC	AAC	GGC	ATC	TGT	AAG	AAC		691
V	G	C	D	F	E	I	D	S	G	A	M	E	D	R	C	G	V	C	H		2073
GTG	GGC	TGT	GAC	TTC	GAG	ATT	GAC	TCC	GGT	GCT	ATG	GAG	GAC	CGC	TGT	GGT	GTG	TGC	CAC		711
G	N	G	S	T	C	H	T	V	S	G	T	F	E	E	A	E	G	L	G		2133
GGC	AAC	GGC	TCC	ACC	TGC	CAC	ACC	GTG	AGC	GGG	ACC	TTC	GAG	GAG	GCC	GAG	GGC	CTG	GGG		731
Y	V	D	V	G	L	I	P	A	G	A	R	E	I	R	I	Q	E	V	A		2193
TAT	GTG	GAT	GTG	GGG	CTG	ATC	CCA	GCC	GGC	GCA	CGC	GAG	ATC	CGC	ATC	CAA	GAG	GTT	GCC		751
E	A	A	N	F	L	A	L	R	S	E	D	P	E	K	Y	F	L	N	G		2253
GAG	GCT	GCC	AAC	TTC	CTG	GCA	CTG	CGG	AGT	GAG	GAC	CCG	GAG	AAG	TAC	TTC	CTC	AAT	GGT		771
G	W	T	I	Q	W	N	G	D	Y	Q	V	A	G	T	T	F	T	Y	A		2313
GGC	TGG	ACC	ATC	CAG	TGG	AAC	GGG	GAC	TAC	CAG	GTG	GCA	GGG	ACC	ACC	TTC	ACA	TAC	GCA		791
R	R	G	N	W	E	N	L	T	S	P	G	P	T	K	E	P	V	W	I		2373
CGC	AGG	GGC	AAC	TGG	GAG	AAC	CTC	ACG	TCC	CCG	GGT	CCC	ACC	AAG	GAG	CCT	GTC	TGG	ATC		811
Q	L	L	F	Q	E	S	N	P	G	V	H	Y	E	Y	T	I	H	R	E		2433
CAG	CTG	CTG	TTC	CAG	GAG	AGC	AAC	CCT	GGG	GTG	CAC	TAC	GAG	TAC	ACC	ATC	CAC	AGG	GAG		831
A	G	G	H	D	E	V	P	P	P	V	F	S	W	H	Y	G	P	W	T		2493
GCA	GGT	GGC	CAC	GAC	GAG	GTC	CCG	CCG	CCC	GTG	TTC	TCC	TGG	CAT	TAT	GGG	CCC	TGG	ACC		851
K	C	T	V	T	C	G	R	G	V	Q	R	Q	N	V	Y	C	L	E	R		

FIG. 9B.

AAG	TGC	ACA	GTC	ACC	TGC	GGC	AGA	GGT	GTG	CAG	AGG	CAG	AAT	GTG	TAC	TGC	TTG	GAG	CGG	2553
Q	A	G	P	V	D	E	E	H	C	D	P	L	G	R	P	D	D	Q	Q	871
CAG	GCA	GGG	CCC	GTG	GAC	GAG	GAG	CAC	TGT	GAC	CCC	CTG	GGC	CGG	CCT	GAT	GAC	CAA	CAG	2613
R	K	C	S	E	Q	P	C	P	A	R	W	W	A	G	E	W	Q	L	C	891
AGG	AAG	TGC	AGC	GAG	CAG	CCC	TGC	CCT	GCC	AGG	TGG	TGG	GCA	GGT	GAG	TGG	CAG	CTG	TGC	2673
S	S	S	C	G	P	G	G	L	S	R	R	A	V	L	C	I	R	S	V	911
TCC	AGC	TCC	TGC	GGG	CCT	GGG	GGC	CTC	TCC	CGC	CGG	GCC	GTG	CTC	TGC	ATC	CGC	AGC	GTG	2733
G	L	D	E	Q	S	A	L	E	P	P	A	C	E	H	L	P	R	P	P	931
GGG	CTG	GAT	GAG	CAG	AGC	GCC	CTG	GAG	CCA	CCC	GCC	TGT	GAA	CAC	CTT	CCC	CGG	CCC	CCT	2793
T	E	T	P	C	N	R	H	V	P	C	P	A	T	W	A	V	G	N	W	951
ACT	GAA	ACC	CCT	TGC	AAC	CGC	CAT	GTA	CCC	TGT	CCG	GCC	ACC	TGG	GCT	GTG	GGG	AAC	TGG	2853
S	Q	C	S	V	T	C	G	E	G	T	Q	R	R	N	V	L	C	T	N	971
TCT	CAG	TGC	TCA	GTG	ACA	TGT	GGG	GAG	GGC	ACT	CAG	CGC	CGA	AAT	GTG	CTC	TGC	ACC	AAT	2913
D	T	G	V	P	C	D	E	A	Q	Q	P	A	S	E	V	T	C	S	L	991
GAC	ACC	GGT	GTC	CCC	TGT	GAC	GAG	GCC	CAG	CAG	CCA	GCC	AGC	GAA	GTC	ACC	TGC	TCT	CTG	2973
P	L	C	R	W	P	L	G	T	L	G	P	E	G	S	G	S	G	S	S	1011
CCA	CTC	TGT	CGG	TGG	CCC	CTG	GGC	ACA	CTG	GGC	CCT	GAA	GGC	TCA	GGC	AGC	GGC	TCC	TCC	3033
S	H	E	L	F	N	E	A	D	P	I	P	H	H	L	A	P	R	P	S	1031
AGC	CAC	GAG	CTC	TTC	AAC	GAG	GCT	GAC	TTC	ATC	CCG	CAC	CAC	CTG	GCC	CCA	CGC	CCT	TCA	3093
P	A	S	S	P	K	P	G	T	M	G	N	A	I	E	E	E	A	P	E	1051
CCC	GCC	TCA	TCA	CCC	AAG	CCA	GGC	ACC	ATG	GGC	AAC	GCC	ATT	GAG	GAG	GAG	GCT	CCA	GAG	3153
L	D	L	P	G	P	V	F	V	D	D	F	Y	Y	D	Y	N	F	I	N	1071
CTG	GAC	CTG	CCG	GGG	CCC	GTG	TTT	GTG	GAC	GAC	TTC	TAC	TAC	GAC	TAC	AAT	TTC	ATC	AAT	3213
F	H	E	D	L	S	Y	G	P	S	E	E	P	D	L	D	L	A	G	T	1091
TTC	CAC	GAG	GAT	CTG	TCC	TAC	GGG	CCC	TCT	GAG	GAG	CCC	GAT	CTA	GAC	CTG	GCG	GGG	ACA	3273
G	D	R	T	P	P	P	H	S	R	P	A	A	P	S	T	G	S	P	V	1111
GGG	GAC	CGG	ACA	CCC	CCA	CCA	CAC	AGC	CGT	CCT	GCT	GCG	CCC	TCC	ACG	GGT	AGC	CCT	GTG	3333
P	A	T	E	P	P	A	A	K	E	E	G	V	L	G	P	W	S	P	S	1131
CCT	GCC	ACA	GAG	CCT	CCT	GCA	GCC	AAG	GAG	GAG	GGG	GTA	CTG	GGA	CCT	TGG	TCC	CCG	AGC	3393
P	W	P	S	Q	A	G	R	S	P	P	P	P	S	E	Q	T	P	G	N	1151
CCT	TGG	CCT	AGC	CAG	GCC	GGC	CGC	TCC	CCA	CCC	CCA	CCC	TCA	GAG	CAG	ACC	CCT	GGG	AAC	3453
P	L	I	N	F	L	P	E	K	D	T	P	I	G	A	P	D	L	G	L	1171
CCT	TTG	ATC	AAT	TTC	CTG	CCT	GAG	GAA	GAC	ACC	CCC	ATA	GGG	GCC	CCA	GAT	CTT	GGG	CTC	3513
P	S	L	S	W	P	R	V	S	T	D	G	L	Q	T	P	A	T	P	E	1191
CCC	AGC	CTG	TCC	TGG	CCC	AGG	GTT	TCC	ACT	GAT	GGC	CTG	CAG	ACA	CCT	GCC	ACC	CCT	GAG	3573
S	Q	N	D	F	P	V	G	K	D	S	Q	S	Q	L	P	P	P	W	R	1211
AGC	CAA	AAT	GAT	TTC	CCA	GTT	GGC	AAG	GAC	AGC	CAG	AGC	CAG	CTG	CCC	CCT	CCA	TGG	CGG	3633
D	R	T	N	E	V	F	K	D	D	E	E	P	K	G	R	G	A	P	H	1231
GAC	AGG	ACC	AAT	GAG	GTT	TTC	AAG	GAT	GAT	GAG	GAA	CCC	AAG	GGC	CGC	GGA	GCA	CCC	CAC	3693
L	P	P	R	P	S	S	T	L	P	P	L	S	P	V	G	S	T	H	S	1251
CTG	CCC	CCG	AGA	CCC	AGC	TCC	ACG	CTG	CCC	CCT	TTG	TCC	CCT	GTT	GGC	AGC	ACC	CAC	TCC	3753
S	P	S	P	D	V	A	E	L	W	T	G	G	T	V	A	W	E	P	A	1271
TCT	CCT	AGT	CCT	GAC	GTG	GCG	GAG	CTG	TGG	ACA	GGA	GGC	ACA	GTG	GCC	TGG	GAG	CCA	GCT	3813

FIG. 9D.

AAC	GCC	AGC	TGG	CAA	GCG	GGA	AAC	TGG	AGC	GAG	TGC	TCT	ACC	ACC	TGT	GGC	CTG	GGT	GCG	4293
V	W	R	P	V	R	C	S	S	G	R	D	E	D	C	A	P	A	G	R	1451
GTC	TGG	AGG	CCG	GTG	CGC	TGT	AGC	TCC	GGC	CGG	GAT	GAG	GAC	TGC	GCC	CCC	GCT	GGC	CGG	4353
P	Q	P	A	R	R	C	H	L	R	P	C	A	T	W	H	S	G	N	W	1471
CCC	CAG	CCT	GCC	CGC	CGC	TGC	CAC	CTG	CGG	CCC	TGT	GCC	ACC	TGG	CAC	TCA	GGC	AAC	TGG	4413
S	K	C	S	R	S	C	G	G	G	S	S	V	R	D	V	Q	C	V	D	1491
AGT	AAG	TGC	TCC	CGC	AGC	TGC	GGC	GGA	GGT	TCC	TCA	GTG	CGG	GAC	GTG	CAG	TGT	GTG	GAC	4473
T	R	D	L	R	P	L	R	P	F	H	C	Q	P	G	P	A	K	P	P	1511
ACA	CGG	GAC	CTC	CGG	CCA	CTG	CGG	CCC	TTC	CAT	TGT	CAG	CCC	GGG	CCT	GCC	AAG	CCG	CCT	4533
A	H	R	P	C	G	A	Q	P	C	L	S	W	Y	T	S	S	W	R	E	1531
GCG	CAC	CGG	CCC	TGC	GGG	GCC	CAG	CCC	TGC	CTC	AGC	TGG	TAC	ACA	TCT	TCC	TGG	AGG	GAG	4593
C	S	E	A	C	G	G	G	E	Q	Q	R	L	V	T	C	P	E	P	G	1551
TGC	TCC	GAG	GCC	TGT	GGC	GGT	GGT	GAG	CAG	CAG	CGT	CTA	GTG	ACC	TGC	CCG	GAG	CCA	GGC	4653
L	C	E	E	A	L	R	P	N	T	T	R	P	C	N	T	H	P	C	T	1571
CTC	TGC	GAG	GAG	GCG	CTG	AGA	CCC	AAC	ACC	ACC	CGG	CCC	TGC	AAC	ACC	CAC	CCC	TGC	ACG	4713
	V	V	A																	
Q	W	V	V	G	P	W	G	Q	C	S	A	P	C	G	G	G	V	Q	R	1591
CAG	TGG	GTG	GTG	GGG	CCC	TGG	GGC	CAG	TGC	TCA	GCC	CCC	TGT	GGT	GGT	GGT	GTG	CAG	CGG	4773
R	L	V	K	C	V	N	T	Q	T	G	L	P	E	E	D	S	D	Q	C	1611
CGC	CTG	GTC	AAG	TGT	GTG	AAC	ACC	CAG	ACA	GGG	CTG	CCC	GAG	GAA	GAC	AGT	GAC	CAG	TGT	4833
G	H	E	A	W	P	E	S	S	R	P	C	G	T	E	D	C	E	P	V	1631
GGC	CAC	GAG	GCC	TGG	CCT	GAG	AGC	TCC	CGG	CCG	TGT	GGC	ACC	GAG	GAT	TGT	GAG	CCC	GTC	4893
E	P	P	R	C	E	R	D	R	L	S	F	G	F	C	E	T	L	R	L	1651
GAG	CCT	CCC	CGC	TGT	GAG	CGG	GAC	CGC	CTG	TCC	TTC	GGG	TTC	TGC	GAG	ACG	CTG	CGC	CTA	4953
L	G	R	C	Q	L	P	T	I	R	T	Q	C	C	R	S	C	S	P	P	1671
CTG	GGC	CGC	TGC	CAG	CTG	CCC	ACC	ATC	CGC	ACC	CAG	TGC	TGC	CGC	TCG	TGC	TCT	CCG	CCC	5013
S	H	G	A	P	S	R	G	H	Q	R	V	A	R	R	*					1687
AGC	CAC	GGC	GCC	CCC	TCC	CGA	GGC	CAT	CAG	CGG	GTT	GCC	CGC	CGC	TGA					5061

CTGTGCCAGGATGCACAGACCGACCGACAGACCTCAGTGCCCAACACGGGCTGTGGCGGAGCTCCCGCCCCCTGCGCCC

FIG. 9E.

TAATGGTGCTAACCCCTCTCACTACCCAGCAGCAGGCTGGGGACCTCCTCCCCCTCAAAAAAGGTATTTTTTTATTCT
AACAGTTTGTGTAACTTTATTATGATTTTACATAAATGAGCATCTACCAAAAAAAAAAAAAAGGGCGGCCGCTAGAC
TAGTCTAGAGAAAAACCTC

FIG. 9F.

SEQUENCE LISTING

<110> Kapeller-Libermann, Rosana

<120> 27875, A Novel Human ADAM-TS Homolog

<130> 5800-39-1-PC

<150> 09/426,282

<151> 1999-10-25

<160> 13

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 1686

<212> PRT

<213> Homo sapiens

<220>

<221> VARIANT

<222> (1)...(1686)

<223> Xaa = Any Amino Acid

<400> 1

```

Met Pro Gly Gly Pro Ser Pro Arg Ser Pro Ala Pro Leu Leu Arg Pro
1      5      10      15
Leu Leu Leu Leu Cys Ala Leu Ala Pro Gly Ala Pro Gly Pro Ala
20      25      30
Pro Gly Arg Ala Thr Glu Gly Arg Ala Ala Leu Asp Ile Val His Pro
35      40      45
Val Arg Val Asp Ala Gly Gly Ser Phe Leu Ser Tyr Glu Leu Trp Pro
50      55      60
Arg Ala Leu Arg Lys Arg Asp Val Ser Val Arg Arg Asp Ala Pro Ala
65      70      75      80
Phe Tyr Glu Leu Gln Tyr Arg Gly Arg Glu Leu Arg Phe Asn Leu Thr
85      90      95
Ala Asn Gln His Leu Leu Ala Pro Gly Phe Val Ser Glu Thr Arg Arg
100     105     110
Arg Gly Gly Leu Gly Arg Ala His Ile Arg Ala His Thr Pro Ala Cys
115     120     125
His Leu Leu Gly Glu Val Gln Asp Pro Glu Leu Glu Gly Gly Leu Ala
130     135     140
Ala Ile Ser Ala Cys Asp Gly Leu Lys Gly Val Phe Gln Leu Ser Asn
145     150     155     160
Glu Asp Tyr Phe Ile Glu Pro Leu Asp Ser Ala Pro Ala Arg Pro Gly
165     170     175
His Ala Gln Pro His Val Val Tyr Lys Arg Gln Ala Pro Glu Arg Leu
180     185     190
Ala Gln Arg Gly Asp Ser Ser Ala Pro Ser Thr Cys Gly Val Gln Val
195     200     205
Tyr Pro Glu Leu Glu Ser Arg Arg Glu Arg Trp Glu Gln Arg Gln Gln
210     215     220
Trp Arg Arg Pro Arg Leu Arg Arg Leu His Gln Arg Ser Val Ser Lys
225     230     235     240
Glu Lys Trp Val Glu Thr Leu Val Val Ala Asp Ala Lys Met Val Glu
245     250     255
Tyr His Gly Gln Pro Gln Val Glu Ser Tyr Val Leu Thr Ile Met Asn
260     265     270
Met Val Ala Gly Leu Phe His Asp Pro Ser Ile Gly Asn Pro Ile His
275     280     285
Ile Thr Ile Val Arg Leu Val Leu Leu Glu Asp Glu Glu Glu Asp Leu
290     295     300
Lys Ile Thr His His Ala Asp Asn Thr Leu Lys Ser Phe Cys Lys Trp

```

305		310		315		320
Gln Lys Ser Ile	Asn Met Lys Gly Asp	Ala His Pro Leu His His Asp				
	325	330	335			
Thr Ala Ile Leu	Leu Thr Arg Lys Asp	Leu Cys Ala Ala Met Asn Arg				
	340	345	350			
Pro Cys Glu Thr	Leu Gly Leu Ser His Val Ala Gly	Met Cys Gln Pro				
	355	360	365			
His Arg Ser Cys	Ser Ile Asn Glu Asp Thr Gly	Leu Pro Leu Ala Phe				
	370	375	380			
Thr Val Ala His	Glu Leu Gly His Ser Phe Gly	Ile Gln His Asp Gly				
	385	390	395			
Ser Gly Asn Asp	Cys Glu Pro Val Gly Lys Arg	Pro Phe Ile Met Ser				
	405	410	415			
Pro Gln Leu Leu	Tyr Asp Ala Ala Pro Leu Thr Trp	Ser Arg Cys Ser				
	420	425	430			
Arg Gln Tyr Ile	Thr Arg Phe Leu Asp Arg Gly Trp	Gly Leu Cys Leu				
	435	440	445			
Asp Asp Pro Pro	Ala Lys Asp Ile Ile Asp Phe	Pro Ser Val Pro Pro				
	450	455	460			
Gly Val Leu Tyr	Asp Val Ser His Gln Cys Arg	Leu Gln Tyr Gly Ala				
	465	470	475			
Tyr Ser Ala Phe	Cys Glu Asp Met Asp Asn Val	Cys His Thr Leu Trp				
	485	490	495			
Cys Ser Val Gly	Thr Thr Cys His Ser Lys Leu Asp	Ala Ala Val Asp				
	500	505	510			
Gly Thr Arg Cys	Gly Glu Asn Lys Trp Cys Leu Ser	Gly Glu Cys Val				
	515	520	525			
Pro Val Gly Phe	Arg Pro Glu Ala Val Asp Gly	Gly Trp Ser Gly Trp				
	530	535	540			
Ser Ala Trp Ser	Ile Cys Ser Arg Ser Cys Gly	Met Gly Val Gln Ser				
	545	550	555			
Ala Glu Arg Gln	Cys Thr Gln Pro Thr Pro Lys Tyr	Lys Gly Arg Tyr				
	565	570	575			
Cys Val Gly Glu	Arg Lys Arg Phe Arg Leu Cys Asn	Leu Gln Ala Cys				
	580	585	590			
Pro Ala Gly Xaa	Pro Ser Phe Arg His Val Gln Cys	Ser His Phe Asp				
	595	600	605			
Ala Met Leu Tyr	Lys Gly Gln Leu His Thr Trp	Val Pro Val Val Asn				
	610	615	620			
Asp Val Asn Pro	Cys Glu Leu His Cys Arg Pro Ala	Asn Glu Tyr Phe				
	625	630	635			
Ala Glu Lys Leu	Arg Asp Ala Val Val Asp Gly	Thr Pro Cys Tyr Gln				
	645	650	655			
Val Arg Ala Ser	Arg Asp Leu Cys Ile Asn Gly Ile	Cys Lys Asn Val				
	660	665	670			
Gly Cys Asp Phe	Glu Ile Asp Ser Gly Ala Met Glu	Asp Arg Cys Gly				
	675	680	685			
Val Cys His Gly	Asn Gly Ser Thr Cys His Thr Val	Ser Gly Thr Phe				
	690	695	700			
Glu Glu Ala Glu	Gly Leu Gly Tyr Val Asp Val Gly	Leu Ile Pro Ala				
	705	710	715			
Gly Ala Arg Glu	Ile Arg Ile Gln Glu Val Ala Glu	Ala Ala Asn Phe				
	725	730	735			
Leu Ala Leu Arg	Ser Glu Asp Pro Glu Lys Tyr Phe	Leu Asn Gly Gly				
	740	745	750			
Trp Thr Ile Gln	Trp Asn Gly Asp Tyr Gln Val Ala	Gly Thr Thr Phe				
	755	760	765			
Thr Tyr Ala Arg	Arg Gly Asn Trp Glu Asn Leu Thr	Ser Pro Gly Pro				
	770	775	780			
Thr Lys Glu Pro	Val Trp Ile Gln Leu Leu Phe Gln	Glu Ser Asn Pro				
	785	790	795			
Gly Val His Tyr	Glu Tyr Thr Ile His Arg Glu Ala	Gly Gly His Asp				
	805	810	815			
Glu Val Pro Pro	Pro Val Phe Ser Trp His Tyr Gly	Pro Trp Thr Lys				
	820	825	830			
Cys Thr Val Thr	Cys Gly Arg Gly Val Gln Arg Gln	Asn Val Tyr Cys				
	835	840	845			

Leu Glu Arg Gln Ala Gly Pro Val Asp Glu Glu His Cys Asp Pro Leu
 850 855 860
 Gly Arg Pro Asp Asp Gln Gln Arg Lys Cys Ser Glu Gln Pro Cys Pro
 865 870 875 880
 Ala Arg Trp Trp Ala Gly Glu Trp Gln Leu Cys Ser Ser Ser Cys Gly
 885 890 895
 Pro Gly Gly Leu Ser Arg Arg Ala Val Leu Cys Ile Arg Ser Val Gly
 900 905 910
 Leu Asp Glu Gln Ser Ala Leu Glu Pro Pro Ala Cys Glu His Leu Pro
 915 920 925
 Arg Pro Pro Thr Glu Thr Pro Cys Asn Arg His Val Pro Cys Pro Ala
 930 935 940
 Thr Trp Ala Val Gly Asn Trp Ser Gln Cys Ser Val Thr Cys Gly Glu
 945 950 955 960
 Gly Thr Gln Arg Arg Asn Val Leu Cys Thr Asn Asp Thr Gly Val Pro
 965 970 975
 Cys Asp Glu Ala Gln Gln Pro Ala Ser Glu Val Thr Cys Ser Leu Pro
 980 985 990
 Leu Cys Arg Trp Pro Leu Gly Thr Leu Gly Pro Glu Gly Ser Gly Ser
 995 1000 1005
 Gly Ser Ser Ser His Glu Leu Phe Asn Glu Ala Asp Phe Ile Pro His
 1010 1015 1020
 His Leu Ala Pro Arg Pro Ser Pro Ala Ser Ser Pro Lys Pro Gly Thr
 1025 1030 1035 1040
 Met Gly Asn Ala Ile Glu Glu Glu Ala Pro Glu Leu Asp Leu Pro Gly
 1045 1050 1055
 Pro Val Phe Val Asp Asp Phe Tyr Tyr Asp Tyr Asn Phe Ile Asn Phe
 1060 1065 1070
 His Glu Asp Leu Ser Tyr Gly Pro Ser Glu Glu Pro Asp Leu Asp Leu
 1075 1080 1085
 Ala Gly Thr Gly Asp Arg Thr Pro Pro Pro His Ser Arg Pro Ala Ala
 1090 1095 1100
 Pro Ser Thr Gly Ser Pro Val Pro Ala Thr Glu Pro Pro Ala Ala Lys
 1105 1110 1115 1120
 Glu Glu Gly Val Leu Gly Pro Trp Ser Pro Ser Pro Trp Pro Ser Gln
 1125 1130 1135
 Ala Gly Arg Ser Pro Pro Pro Pro Ser Glu Gln Thr Pro Gly Asn Pro
 1140 1145 1150
 Leu Ile Asn Phe Leu Pro Glu Glu Asp Thr Pro Ile Gly Ala Pro Asp
 1155 1160 1165
 Leu Gly Leu Pro Ser Leu Ser Trp Pro Arg Val Ser Thr Asp Gly Leu
 1170 1175 1180
 Gln Thr Pro Ala Thr Pro Glu Ser Gln Asn Asp Phe Pro Val Gly Lys
 1185 1190 1195 1200
 Asp Ser Gln Ser Gln Leu Pro Pro Pro Trp Arg Asp Arg Thr Asn Glu
 1205 1210 1215
 Val Phe Lys Asp Asp Glu Glu Pro Lys Gly Arg Gly Ala Pro His Leu
 1220 1225 1230
 Pro Pro Arg Pro Ser Ser Thr Leu Pro Pro Leu Ser Pro Val Gly Ser
 1235 1240 1245
 Thr His Ser Ser Pro Ser Pro Asp Val Ala Glu Leu Trp Thr Gly Gly
 1250 1255 1260
 Thr Val Ala Trp Glu Pro Ala Leu Glu Gly Gly Leu Gly Pro Val Asp
 1265 1270 1275 1280
 Ser Glu Leu Trp Pro Thr Val Gly Val Ala Ser Leu Leu Pro Pro Pro
 1285 1290 1295
 Ile Ala Pro Leu Pro Glu Met Lys Val Arg Asp Ser Ser Leu Glu Pro
 1300 1305 1310
 Gly Thr Pro Ser Phe Pro Ala Pro Gly Pro Gly Ser Trp Asp Leu Gln
 1315 1320 1325
 Thr Val Ala Val Trp Gly Thr Phe Leu Pro Thr Thr Leu Thr Gly Leu
 1330 1335 1340
 Gly His Met Pro Glu Pro Ala Leu Asn Pro Gly Pro Lys Gly Gln Pro
 1345 1350 1355 1360
 Glu Ser Leu Thr Pro Glu Val Pro Leu Ser Ser Arg Leu Leu Ser Thr
 1365 1370 1375
 Pro Ala Trp Asp Ser Pro Ala Asn Ser His Arg Val Pro Glu Thr Gln

```

      1380      1385      1390
Pro Leu Ala Pro Ser Leu Ala Glu Ala Gly Pro Pro Ala Asp Pro Leu
      1395      1400      1405
Val Val Arg Asn Ala Ser Trp Gln Ala Gly Asn Trp Ser Glu Cys Ser
      1410      1415      1420
Thr Thr Cys Gly Leu Gly Ala Val Trp Arg Pro Val Arg Cys Ser Ser
1425      1430      1435      1440
Gly Arg Asp Glu Asp Cys Ala Pro Ala Gly Arg Pro Gln Pro Ala Arg
      1445      1450      1455
Arg Cys His Leu Arg Pro Cys Ala Thr Trp His Ser Gly Asn Trp Ser
      1460      1465      1470
Lys Cys Ser Arg Ser Cys Asp Gly Gly Ser Ser Val Arg Asp Val Gln
      1475      1480      1485
Cys Val Asp Thr Arg Asp Leu Arg Pro Leu Arg Pro Phe His Cys Gln
      1490      1495      1500
Pro Gly Pro Ala Lys Pro His Ala His Arg Pro Cys Gly Ala Gln Pro
1505      1510      1515      1520
Cys Leu Ser Trp Tyr Thr Ser Ser Trp Arg Glu Cys Ser Glu Ala Cys
      1525      1530      1535
Gly Gly Gly Glu Gln Gln Arg Leu Val Thr Cys Pro Glu Pro Gly Leu
      1540      1545      1550
Cys Glu Glu Ala Leu Arg Pro Asn Thr Thr Arg Pro Cys Asn Thr His
      1555      1560      1565
Pro Cys Thr Gln Trp Val Val Gly Pro Trp Gly Gln Cys Ser Ala Pro
      1570      1575      1580
Cys Gly Gly Gly Val Gln Arg Arg Leu Val Lys Cys Val Asn Thr Gln
1585      1590      1595      1600
Thr Gly Leu Pro Glu Glu Asp Ser Asp Gln Cys Gly His Glu Ala Trp
      1605      1610      1615
Pro Glu Ser Ser Arg Pro Cys Gly Thr Glu Asp Cys Glu Pro Val Glu
      1620      1625      1630
Pro Pro Arg Cys Glu Arg Asp Arg Leu Ser Phe Gly Phe Cys Glu Thr
      1635      1640      1645
Leu Arg Leu Leu Gly Arg Cys Gln Leu Pro Thr Ile Arg Thr Gln Cys
      1650      1655      1660
Cys Arg Ser Cys Ser Pro Ser His Gly Ala Pro Ser Arg Gly His
1665      1670      1675      1680
Gln Arg Val Ala Arg Arg
      1685

```

<210> 2
 <211> 5353
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> (36)...(5093)

<400> 2
 aaagggaata agcttgccggc cgcccgggttc ctgcc atg ccc ggc ggc ccc agt
 53

Met Pro Gly Gly Pro Ser
 1 5

ccc cgc agc ccc gcg cct ttg ctg cgc ccc ctc ctc ctg ctc ctc tgc
 101

Pro Arg Ser Pro Ala Pro Leu Leu Arg Pro Leu Leu Leu Leu Cys
 10 15 20

gct ctg gct ccc ggc gcc ccc gga ccc gca cca gga cgt gca acc gag
 149

Ala Leu Ala Pro Gly Ala Pro Gly Pro Ala Pro Gly Arg Ala Thr Glu
 25 30 35

ggc cgg gcg gca ctg gac atc gtg cac ccg gtt cga gtc gac gcg ggg
 197

Gly Arg Ala Ala Leu Asp Ile Val His Pro Val Arg Val Asp Ala Gly
 40 45 50
 ggc tcc ttc ctg tcc tac gag ctg tgg ccc cgc gca ctg cgc aag cgg
 245
 Gly Ser Phe Leu Ser Tyr Glu Leu Trp Pro Arg Ala Leu Arg Lys Arg
 55 60 65 70
 gat gta tct gtg cgc cga gac gcg ccc gcc ttc tac gag cta caa tac
 293
 Asp Val Ser Val Arg Arg Asp Ala Pro Ala Phe Tyr Glu Leu Gln Tyr
 75 80 85
 cgc ggg cgc gag ctg cgc ttc aac ctg acc gcc aat cag cac ctg ctg
 341
 Arg Gly Arg Glu Leu Arg Phe Asn Leu Thr Ala Asn Gln His Leu Leu
 90 95 100
 gcg ccc ggc ttt gtg agc gag acg cgg cgg cgc ggc ggc ctg ggc cgc
 389
 Ala Pro Gly Phe Val Ser Glu Thr Arg Arg Arg Gly Gly Leu Gly Arg
 105 110 115
 gcg cac atc cgg gcc cac acc ccg gcc tgc cac ctg ctt ggc gag gtg
 437
 Ala His Ile Arg Ala His Thr Pro Ala Cys His Leu Leu Gly Glu Val
 120 125 130
 cag gac cct gag ctc gag ggt ggc ctg gcg gcc atc agc gcc tgc gac
 485
 Gln Asp Pro Glu Leu Glu Gly Gly Leu Ala Ala Ile Ser Ala Cys Asp
 135 140 145 150
 ggc ctg aaa ggt gtg ttc caa ctc tcc aac gag gac tac ttc att gag
 533
 Gly Leu Lys Gly Val Phe Gln Leu Ser Asn Glu Asp Tyr Phe Ile Glu
 155 160 165
 ccc ctg gac agt gcc ccg gcc cgg cct ggc cac gcc cag ccc cat gtg
 581
 Pro Leu Asp Ser Ala Pro Ala Arg Pro Gly His Ala Gln Pro His Val
 170 175 180
 gtg tac aag cgt cag gcc ccg gag agg ctg gca cag cgg ggt gat tcc
 629
 Val Tyr Lys Arg Gln Ala Pro Glu Arg Leu Ala Gln Arg Gly Asp Ser
 185 190 195
 agt gct cca agc acc tgt gga gtg caa gtg tac cca gag ctg gag tct
 677
 Ser Ala Pro Ser Thr Cys Gly Val Gln Val Tyr Pro Glu Leu Glu Ser
 200 205 210
 cga cgg gag cgt tgg gag cag cgg cag cag tgg cgg cgg cca cgg ctg
 725
 Arg Arg Glu Arg Trp Glu Gln Arg Gln Gln Trp Arg Arg Pro Arg Leu
 215 220 225 230
 agg cgt cta cac cag cgg tcg gtc agc aaa gag aag tgg gtg gag acc
 773
 Arg Arg Leu His Gln Arg Ser Val Ser Lys Glu Lys Trp Val Glu Thr
 235 240 245
 ctg gta gta gct gat gcc aaa atg gtg gag tac cac gga cag ccg cag
 821
 Leu Val Val Ala Asp Ala Lys Met Val Glu Tyr His Gly Gln Pro Gln
 250 255 260

gtt gag agc tat gtg ctg acc atc atg aac atg gtg gct ggc ctg ttt
 869
 Val Glu Ser Tyr Val Leu Thr Ile Met Asn Met Val Ala Gly Leu Phe
 265 270 275

cat gac ccc agc att ggg aac ccc atc cac atc acc att gtg cgc ctg
 917
 His Asp Pro Ser Ile Gly Asn Pro Ile His Ile Thr Ile Val Arg Leu
 280 285 290

gtc ctg ctg gaa gat gag gag gag gac cta aag atc acg cac cat gca
 965
 Val Leu Leu Glu Asp Glu Glu Glu Asp Leu Lys Ile Thr His His Ala
 295 300 305 310

gac aac acc ctg aag agc ttc tgc aag tgg cag aaa agc atc aac atg
 1013
 Asp Asn Thr Leu Lys Ser Phe Cys Lys Trp Gln Lys Ser Ile Asn Met
 315 320 325

aag ggg gat gcc cat ccc ctg cac cat gac act gcc atc ctg ctc acc
 1061
 Lys Gly Asp Ala His Pro Leu His His Asp Thr Ala Ile Leu Leu Thr
 330 335 340

aga aag gac ctg tgt gca gcc atg aac cgg ccc tgt gag acc ctg gga
 1109
 Arg Lys Asp Leu Cys Ala Ala Met Asn Arg Pro Cys Glu Thr Leu Gly
 345 350 355

ctg tcc cat gtg gcg ggc atg tgc cag ccg cac cgc agc tgc agc atc
 1157
 Leu Ser His Val Ala Gly Met Cys Gln Pro His Arg Ser Cys Ser Ile
 360 365 370

aac gag gac acg ggc ctg ccg ctg gcc ttc act gta gcc cac gag ctc
 1205
 Asn Glu Asp Thr Gly Leu Pro Leu Ala Phe Thr Val Ala His Glu Leu
 375 380 385 390

ggg cac agt ttt ggc att cag cat gac gga agc ggc aat gac tgt gag
 1253
 Gly His Ser Phe Gly Ile Gln His Asp Gly Ser Gly Asn Asp Cys Glu
 395 400 405

ccc gtt ggg aaa cga cct ttc atc atg tct cca cag ctc ctg tac gac
 1301
 Pro Val Gly Lys Arg Pro Phe Ile Met Ser Pro Gln Leu Leu Tyr Asp
 410 415 420

gcc gct ccc ctc acc tgg tcc cgc tgc agc cgc cag tat atc acc agg
 1349
 Ala Ala Pro Leu Thr Trp Ser Arg Cys Ser Arg Gln Tyr Ile Thr Arg
 425 430 435

ttc ctt gac cgt ggg tgg ggc ctg tgc ctg gac gac cct cct gcc aag
 1397
 Phe Leu Asp Arg Gly Trp Gly Leu Cys Leu Asp Asp Pro Pro Ala Lys
 440 445 450

gac att atc gac ttc ccc tcg gtg cca cct ggc gtc ctc tat gat gta
 1445
 Asp Ile Ile Asp Phe Pro Ser Val Pro Pro Gly Val Leu Tyr Asp Val
 455 460 465 470

agc cac cag tgc cgc ctc cag tac ggg gcc tac tct gcc ttc tgc gag
 1493
 Ser His Gln Cys Arg Leu Gln Tyr Gly Ala Tyr Ser Ala Phe Cys Glu
 475 480 485

gac atg gat aat gtc tgc cac aca ctc tgg tgc tct gtg ggg acc acc
 1541
 Asp Met Asp Asn Val Cys His Thr Leu Trp Cys Ser Val Gly Thr Thr
 490 495 500

tgt cac tcc aag ctg gat gca gcc gtg gac ggc acc cgg tgt ggg gag
 1589
 Cys His Ser Lys Leu Asp Ala Ala Val Asp Gly Thr Arg Cys Gly Glu
 505 510 515

aat aag tgg tgt ctc agt ggg gag tgc gta ccc gtg ggc ttc cgg ccc
 1637
 Asn Lys Trp Cys Leu Ser Gly Glu Cys Val Pro Val Gly Phe Arg Pro
 520 525 530

gag gcc gtg gat ggt ggc tgg tct ggc tgg agc gcc tgg tcc atc tgc
 1685
 Glu Ala Val Asp Gly Gly Trp Ser Gly Trp Ser Ala Trp Ser Ile Cys
 535 540 545 550

tca cgg agc tgt ggc atg ggc gta cag agc gcc gag cgg cag tgc acg
 1733
 Ser Arg Ser Cys Gly Met Gly Val Gln Ser Ala Glu Arg Gln Cys Thr
 555 560 565

cag cct acg ccc aaa tac aaa ggc aga tac tgt gtg ggt gag cgc aag
 1781
 Gln Pro Thr Pro Lys Tyr Lys Gly Arg Tyr Cys Val Gly Glu Arg Lys
 570 575 580

cgc ttc cgc ctc tgc aac ctg cag gcc tgc cct gct ggc crc ccc tcc
 1829
 Arg Phe Arg Leu Cys Asn Leu Gln Ala Cys Pro Ala Gly Xaa Pro Ser
 585 590 595

ttc cgc cac gtc cag tgc agc cac ttt gac gct atg ctc tac aag ggc
 1877
 Phe Arg His Val Gln Cys Ser His Phe Asp Ala Met Leu Tyr Lys Gly
 600 605 610

cag ctg cac aca tgg gtg ccc gtg gtc aat gac gtg aac ccc tgc gag
 1925
 Gln Leu His Thr Trp Val Pro Val Val Asn Asp Val Asn Pro Cys Glu
 615 620 625 630

ctg cac tgc cgg ccc gcg aat gag tac ttt gcc gag aag ctg cgg gac
 1973
 Leu His Cys Arg Pro Ala Asn Glu Tyr Phe Ala Glu Lys Leu Arg Asp
 635 640 645

gcc gtg gtc gat ggc acc ccc tgc tac cag gtc cga gcc agc cgg gac
 2021
 Ala Val Val Asp Gly Thr Pro Cys Tyr Gln Val Arg Ala Ser Arg Asp
 650 655 660

ctc tgc atc aac ggc atc tgt aag aac gtg ggc tgt gac ttc gag att
 2069
 Leu Cys Ile Asn Gly Ile Cys Lys Asn Val Gly Cys Asp Phe Glu Ile
 665 670 675

gac tcc ggt gct atg gag gac cgc tgt ggt gtg tgc cac ggc aac ggc
 2117

Asp Ser Gly Ala Met Glu Asp Arg Cys Gly Val Cys His Gly Asn Gly
 680 685 690
 tcc acc tgc cac acc gtg agc ggg acc ttc gag gag gcc gag ggc ctg
 2165
 Ser Thr Cys His Thr Val Ser Gly Thr Phe Glu Glu Ala Glu Gly Leu
 695 700 705 710
 ggg tat gtg gat gtg ggg ctg atc cca gcc ggc gca cgc gag atc cgc
 2213
 Gly Tyr Val Asp Val Gly Leu Ile Pro Ala Gly Ala Arg Glu Ile Arg
 715 720 725
 atc caa gag gtt gcc gag gct gcc aac ttc ctg gca ctg cgg agt gag
 2261
 Ile Gln Glu Val Ala Glu Ala Ala Asn Phe Leu Ala Leu Arg Ser Glu
 730 735 740
 gac ccg gag aag tac ttc ctc aat ggt ggc tgg acc atc cag tgg aac
 2309
 Asp Pro Glu Lys Tyr Phe Leu Asn Gly Gly Trp Thr Ile Gln Trp Asn
 745 750 755
 ggg gac tac cag gtg gca ggg acc acc ttc aca tac gca cgc agg ggc
 2357
 Gly Asp Tyr Gln Val Ala Gly Thr Thr Phe Thr Tyr Ala Arg Arg Gly
 760 765 770
 aac tgg gag aac ctc acg tcc ccg ggt ccc acc aag gag cct gtc tgg
 2405
 Asn Trp Glu Asn Leu Thr Ser Pro Gly Pro Thr Lys Glu Pro Val Trp
 775 780 785 790
 atc cag ctg ctg ttc cag gag agc aac cct ggg gtg cac tac gag tac
 2453
 Ile Gln Leu Leu Phe Gln Glu Ser Asn Pro Gly Val His Tyr Glu Tyr
 795 800 805
 acc atc cac agg gag gca ggt ggc cac gac gag gtc ccg ccg ccc gtg
 2501
 Thr Ile His Arg Glu Ala Gly Gly His Asp Glu Val Pro Pro Pro Val
 810 815 820
 ttc tcc tgg cat tat ggg ccc tgg acc aag tgc aca gtc acc tgc ggc
 2549
 Phe Ser Trp His Tyr Gly Pro Trp Thr Lys Cys Thr Val Thr Cys Gly
 825 830 835
 aga ggt gtg cag agg cag aat gtg tac tgc ttg gag cgg cag gca ggg
 2597
 Arg Gly Val Gln Arg Gln Asn Val Tyr Cys Leu Glu Arg Gln Ala Gly
 840 845 850
 ccc gtg gac gag gag cac tgt gac ccc ctg ggc cgg cct gat gac caa
 2645
 Pro Val Asp Glu Glu His Cys Asp Pro Leu Gly Arg Pro Asp Asp Gln
 855 860 865 870
 cag agg aag tgc agc gag cag ccc tgc cct gcc agg tgg tgg gca ggt
 2693
 Gln Arg Lys Cys Ser Glu Gln Pro Cys Pro Ala Arg Trp Trp Ala Gly
 875 880 885
 gag tgg cag ctg tgc tcc agc tcc tgc ggg cct ggg ggc ctc tcc cgc
 2741
 Glu Trp Gln Leu Cys Ser Ser Ser Cys Gly Pro Gly Gly Leu Ser Arg
 890 895 900

cgg gcc gtg ctc tgc atc cgc agc gtg ggg ctg gat gag cag agc gcc
2789

Arg Ala Val Leu Cys Ile Arg Ser Val Gly Leu Asp Glu Gln Ser Ala
905 910 915

ctg gag cca ccc gcc tgt gaa cac ctt ccc cgg ccc cct act gaa acc
2837

Leu Glu Pro Pro Ala Cys Glu His Leu Pro Arg Pro Pro Thr Glu Thr
920 925 930

cct tgc aac cgc cat gta ccc tgt ccg gcc acc tgg gct gtg ggg aac
2885

Pro Cys Asn Arg His Val Pro Cys Pro Ala Thr Trp Ala Val Gly Asn
935 940 945 950

tgg tct cag tgc tca gtg aca tgt ggg gag ggc act cag cgc cga aat
2933

Trp Ser Gln Cys Ser Val Thr Cys Gly Glu Gly Thr Gln Arg Arg Asn
955 960 965

gtc ctc tgc acc aat gac acc ggt gtc ccc tgt gac gag gcc cag cag
2981

Val Leu Cys Thr Asn Asp Thr Gly Val Pro Cys Asp Glu Ala Gln Gln
970 975 980

cca gcc agc gaa gtc acc tgc tct ctg cca ctc tgt cgg tgg ccc ctg
3029

Pro Ala Ser Glu Val Thr Cys Ser Leu Pro Leu Cys Arg Trp Pro Leu
985 990 995

ggc aca ctg ggc cct gaa ggc tca ggc agc ggc tcc tcc agc cac gag
3077

Gly Thr Leu Gly Pro Glu Gly Ser Gly Ser Gly Ser Ser Ser His Glu
1000 1005 1010

ctc ttc aac gag gct gac ttc atc ccg cac cac ctg gcc cca cgc cct
3125

Leu Phe Asn Glu Ala Asp Phe Ile Pro His His Leu Ala Pro Arg Pro
1015 1020 1025 1030

tca ccc gcc tca tca ccc aag cca ggc acc atg ggc aac gcc att gag
3173

Ser Pro Ala Ser Ser Pro Lys Pro Gly Thr Met Gly Asn Ala Ile Glu
1035 1040 1045

gag gag gct cca gag ctg gac ctg ccg ggg ccc gtg ttt gtg gac gac
3221

Glu Glu Ala Pro Glu Leu Asp Leu Pro Gly Pro Val Phe Val Asp Asp
1050 1055 1060

ttc tac tac gac tac aat ttc atc aat ttc cac gag gat ctg tcc tac
3269

Phe Tyr Tyr Asp Tyr Asn Phe Ile Asn Phe His Glu Asp Leu Ser Tyr
1065 1070 1075

ggg ccc tct gag gag ccc gat cta gac ctg gcg ggg aca ggg gac cgg
3317

Gly Pro Ser Glu Glu Pro Asp Leu Asp Leu Ala Gly Thr Gly Asp Arg
1080 1085 1090

aca ccc cca cca cac agc cgt cct gct gcg ccc tcc acg ggt agc cct
3365

Thr Pro Pro Pro His Ser Arg Pro Ala Ala Pro Ser Thr Gly Ser Pro
1095 1100 1105 1110

gtg cct gcc aca gag cct cct gca gcc aag gag gag ggg gta ctg gga
 3413
 Val Pro Ala Thr Glu Pro Pro Ala Ala Lys Glu Glu Gly Val Leu Gly
 1115 1120 1125

cct tgg tcc ccg agc cct tgg cct agc cag gcc ggc cgc tcc cca ccc
 3461
 Pro Trp Ser Pro Ser Pro Trp Pro Ser Gln Ala Gly Arg Ser Pro Pro
 1130 1135 1140

cca ccc tca gag cag acc cct ggg aac cct ttg atc aat ttc ctg cct
 3509
 Pro Pro Ser Glu Gln Thr Pro Gly Asn Pro Leu Ile Asn Phe Leu Pro
 1145 1150 1155

gag gaa gac acc ccc ata ggg gcc cca gat ctt ggg ctc ccc agc ctg
 3557
 Glu Glu Asp Thr Pro Ile Gly Ala Pro Asp Leu Gly Leu Pro Ser Leu
 1160 1165 1170

tcc tgg ccc agg gtt tcc act gat ggc ctg cag aca cct gcc acc cct
 3605
 Ser Trp Pro Arg Val Ser Thr Asp Gly Leu Gln Thr Pro Ala Thr Pro
 1175 1180 1185 1190

gag agc caa aat gat ttc cca gtt ggc aag gac agc cag agc cag ctg
 3653
 Glu Ser Gln Asn Asp Phe Pro Val Gly Lys Asp Ser Gln Ser Gln Leu
 1195 1200 1205

ccc cct cca tgg cgg gac agg acc aat gag gtt ttc aag gat gat gag
 3701
 Pro Pro Pro Trp Arg Asp Arg Thr Asn Glu Val Phe Lys Asp Asp Glu
 1210 1215 1220

gaa ccc aag ggc cgc gga gca ccc cac ctg ccc ccg aga ccc agc tcc
 3749
 Glu Pro Lys Gly Arg Gly Ala Pro His Leu Pro Pro Arg Pro Ser Ser
 1225 1230 1235

acg ctg ccc cct ttg tcc cct gtt ggc agc acc cac tcc tct cct agt
 3797
 Thr Leu Pro Pro Leu Ser Pro Val Gly Ser Thr His Ser Ser Pro Ser
 1240 1245 1250

cct gac gtg gcg gag ctg tgg aca gga ggc aca gtg gcc tgg gag cca
 3845
 Pro Asp Val Ala Glu Leu Trp Thr Gly Gly Thr Val Ala Trp Glu Pro
 1255 1260 1265 1270

gct ctg gag ggt ggc ctg ggg cct gtg gac agt gaa ctg tgg ccc act
 3893
 Ala Leu Glu Gly Gly Leu Gly Pro Val Asp Ser Glu Leu Trp Pro Thr
 1275 1280 1285

gtt ggg gtg gct tct ctc ctt cct cct ccc ata gcc cct ctg cca gag
 3941
 Val Gly Val Ala Ser Leu Leu Pro Pro Pro Ile Ala Pro Leu Pro Glu
 1290 1295 1300

atg aag gtc agg gac agt tcc ctg gag ccg ggg act ccc tcc ttc cca
 3989
 Met Lys Val Arg Asp Ser Ser Leu Glu Pro Gly Thr Pro Ser Phe Pro
 1305 1310 1315

gcc cca gga cca ggc tca tgg gac ctg cag act gtg gca gtg tgg ggg
 4037

Ala Pro Gly Pro Gly Ser Trp Asp Leu Gln Thr Val Ala Val Trp Gly
 1320 1325 1330

acc ttc ctc ccc aca acc ctg act ggc ctc ggg cac atg cct gag cct
 4085

Thr Phe Leu Pro Thr Thr Leu Thr Gly Leu Gly His Met Pro Glu Pro
 1335 1340 1345 1350

gcc ctg aac cca gga ccc aag ggt cag cct gag tcc ctc acc cct gag
 4133

Ala Leu Asn Pro Gly Pro Lys Gly Gln Pro Glu Ser Leu Thr Pro Glu
 1355 1360 1365

gtg ccc ctg agc tct agg ctg ctg tcc aca cca gct tgg gac agc ccc
 4181

Val Pro Leu Ser Ser Arg Leu Leu Ser Thr Pro Ala Trp Asp Ser Pro
 1370 1375 1380

gcc aac agc cac aga gtc cct gag acc cag ccg ctg gct ccc agc ctg
 4229

Ala Asn Ser His Arg Val Pro Glu Thr Gln Pro Leu Ala Pro Ser Leu
 1385 1390 1395

gct gaa gcg ggg ccc ccc gcg gac ccg ttg gtt gtc agg aac gcc agc
 4277

Ala Glu Ala Gly Pro Pro Ala Asp Pro Leu Val Val Arg Asn Ala Ser
 1400 1405 1410

tgg caa gcg gga aac tgg agc gag tgc tct acc acc tgt ggc ctg ggt
 4325

Trp Gln Ala Gly Asn Trp Ser Glu Cys Ser Thr Thr Cys Gly Leu Gly
 1415 1420 1425 1430

gcg gtc tgg agg ccg gtg cgc tgt agc tcc ggc cgg gat gag gac tgc
 4373

Ala Val Trp Arg Pro Val Arg Cys Ser Ser Gly Arg Asp Glu Asp Cys
 1435 1440 1445

gcc ccc gct ggc cgg ccc cag cct gcc cgc cgc tgc cac cta cgg ccc
 4421

Ala Pro Ala Gly Arg Pro Gln Pro Ala Arg Arg Cys His Leu Arg Pro
 1450 1455 1460

tgt gcc acc tgg cac tca ggc aac tgg agt aag tgc tcc cgc agc tgc
 4469

Cys Ala Thr Trp His Ser Gly Asn Trp Ser Lys Cys Ser Arg Ser Cys
 1465 1470 1475

gac gga ggt tcc tca gtg cgg gac gtg cag tgt gtg gac aca cgg gac
 4517

Asp Gly Gly Ser Ser Val Arg Asp Val Gln Cys Val Asp Thr Arg Asp
 1480 1485 1490

ctc cgg cca ctg cgg ccc ttc cat tgt cag ccc ggg cct gcc aag ccg
 4565

Leu Arg Pro Leu Arg Pro Phe His Cys Gln Pro Gly Pro Ala Lys Pro
 1495 1500 1505 1510

cat gcg cac cgg ccc tgc ggg gcc cag ccc tgc ctc agc tgg tac aca
 4613

His Ala His Arg Pro Cys Gly Ala Gln Pro Cys Leu Ser Trp Tyr Thr
 1515 1520 1525

tct tcc tgg agg gag tgc tcc gag gcc tgt ggc ggt ggt gag cag cag
 4661

Ser Ser Trp Arg Glu Cys Ser Glu Ala Cys Gly Gly Gly Glu Gln Gln
 1530 1535 1540

cgt cta gtg acc tgc ccg gag cca ggc ctc tgc gag gag gcg ctg aga
4709
Arg Leu Val Thr Cys Pro Glu Pro Gly Leu Cys Glu Glu Ala Leu Arg
1545 1550 1555

ccc aac acc acc cgg ccc tgc aac acc cac ccc tgc acg cag tgg gtg
4757
Pro Asn Thr Thr Arg Pro Cys Asn Thr His Pro Cys Thr Gln Trp Val
1560 1565 1570

gtg ggg ccc tgg ggc cag tgc tca gcc ccc tgt ggt ggt ggt gtc cag
4805
Val Gly Pro Trp Gly Gln Cys Ser Ala Pro Cys Gly Gly Gly Val Gln
1575 1580 1585 1590

cgg cgc ctg gtc aag tgt gtc aac acc cag aca ggg ctg ccc gag gaa
4853
Arg Arg Leu Val Lys Cys Val Asn Thr Gln Thr Gly Leu Pro Glu Glu
1595 1600 1605

gac agt gac cag tgt ggc cac gag gcc tgg cct gag agc tcc cgg ccg
4901
Asp Ser Asp Gln Cys Gly His Glu Ala Trp Pro Glu Ser Ser Arg Pro
1610 1615 1620

tgt ggc acc gag gat tgt gag ccc gtc gag cct ccc cgc tgt gag cgg
4949
Cys Gly Thr Glu Asp Cys Glu Pro Val Glu Pro Pro Arg Cys Glu Arg
1625 1630 1635

gac cgc ctg tcc ttc ggg ttc tgc gag acg ctg cgc cta ctg ggc cgc
4997
Asp Arg Leu Ser Phe Gly Phe Cys Glu Thr Leu Arg Leu Leu Gly Arg
1640 1645 1650

tgc cag ctg ccc acc atc cgc acc cag tgc tgc cgc tcg tgc tct ccg
5045
Cys Gln Leu Pro Thr Ile Arg Thr Gln Cys Cys Arg Ser Cys Ser Pro
1655 1660 1665 1670

ccc agc cac ggc gcc ccc tcc cga ggc cat cag cgg gtt gcc cgc cgc
5093
Pro Ser His Gly Ala Pro Ser Arg Gly His Gln Arg Val Ala Arg Arg
1675 1680 1685

tgactgtgcc aggatgcaca gaccgaccga cagacctcag tgcccaccac gggctgtggc
5153
ggagctcccg cccctgcgc cctaattggtg ctaacccct ctcactaccc agcagcaggc
5213
tggggacctc cteccctca aaaaaggat ttttttattc taacagtttg tgtaacattt
5273
attatgattt tacataaatg agcatctacc aaaaaaaaaa aaaaagggcg gccgctagac
5333
tagtctagag aaaaaacctc
5353

<210> 3
<211> 123
<212> PRT
<213> Homo sapiens

<400> 3
Ala Ser Gly Thr Pro Ser Gly Phe Thr Ile Ala His Glu Leu Gly His
1 5 10 15
Ser Phe Gly Ile Gln His Asp Gly Lys Glu Asn Asp Cys Glu Pro Val
20 25 30

Gly Arg His Pro Tyr Ile Met Ser Arg Gln Leu Gln Tyr Asp Pro Thr
 35 40 45
 Pro Leu Thr Trp Ser Lys Cys Ser Glu Glu Tyr Ile Thr Arg Phe Leu
 50 55 60
 Asp Arg Gly Trp Gly Phe Cys Leu Asp Asp Ile Pro Lys Lys Lys Gly
 65 70 75 80
 Leu Lys Ser Lys Val Ile Ala Pro Gly Val Ile Tyr Asp Val His His
 85 90 95
 Gln Cys Gln Leu Gln Tyr Gly Pro Asn Ala Thr Phe Cys Arg Glu Val
 100 105 110
 Glu Asn Val Cys Gln Thr Leu Val Val Ala Pro
 115 120

<210> 4
 <211> 415
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> (2)...(370)

<400> 4
 c gcg tcc ggg act ccc tct ggc ttt aca att gcc cat gag cta gga cac
 49
 Ala Ser Gly Thr Pro Ser Gly Phe Thr Ile Ala His Glu Leu Gly His
 1 5 10 15

agc ttc ggc atc cag cat gat ggg aaa gaa aat gac tgt gag cct gtg
 97
 Ser Phe Gly Ile Gln His Asp Gly Lys Glu Asn Asp Cys Glu Pro Val
 20 25 30

ggc aga cat ccg tac atc atg tcc cgc cag ctc cag tac gat ccc act
 145
 Gly Arg His Pro Tyr Ile Met Ser Arg Gln Leu Gln Tyr Asp Pro Thr
 35 40 45

ccg ctg aca tgg tcc aag tgc agc gag gag tac atc acc cgc ttc ttg
 193
 Pro Leu Thr Trp Ser Lys Cys Ser Glu Glu Tyr Ile Thr Arg Phe Leu
 50 55 60

gac cga ggc tgg ggg ttc tgt ctt gat gac ata cct aaa aag aaa ggc
 241
 Asp Arg Gly Trp Gly Phe Cys Leu Asp Asp Ile Pro Lys Lys Lys Gly
 65 70 75 80

ttg aag tcc aag gtc att gcc ccc gga gtg atc tat gat gtt cac cac
 289
 Leu Lys Ser Lys Val Ile Ala Pro Gly Val Ile Tyr Asp Val His His
 85 90 95

cag tgc cag cta caa tat gga ccc aat gct acc ttc tgc agg gaa gta
 337
 Gln Cys Gln Leu Gln Tyr Gly Pro Asn Ala Thr Phe Cys Arg Glu Val
 100 105 110

gaa aac gtc tgc cag aca ctt gtg gtg gct ccg tgaagggtt ttgtcgtct
 390
 Glu Asn Val Cys Gln Thr Leu Val Val Ala Pro
 115 120

aagctggacg ctgctgcaga tggaa
 415

<210> 5

<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Heparin binding sites in conserved
Thrombospondin-type 1 motif

<221> VARIANT
<222> (2)...(2)
<223> The S at position 2 can also be G.

<221> VARIANT
<222> (1)...(7)
<223> Xaa = Any Amino Acid

<400> 5
Trp Ser Xaa Trp Ser Xaa Trp
1 5

<210> 6
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> Heparin binding site in conserved
Thrombospondin-type 1 motif

<400> 6
Cys Ser Val Thr Cys Gly
1 5

<210> 7
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Heparin binding site in conserved
Thrombospondin-type 1 motif

<400> 7
Trp Gly Pro Trp Gly Pro Trp
1 5

<210> 8
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> Heparin binding site in conserved
Thrombospondin-type 1 motif

<221> VARIANT
<222> (3)...(3)
<223> The R amino acid position 3 can also be K.

<400> 8
Cys Ser Arg Thr Cys Gly
1 5

<210> 9
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Cleavage site in the proprotein domain recognized
by the furin-like protease.

<400> 9
Arg Arg Arg Arg
1

<210> 10
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Cleavage site in the proprotein domain recognized
by the furin-like protease.

<400> 10
Arg Lys Lys Arg
1

<210> 11
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Consensus sequence for the furin-like protease
cleavage domain.

<221> VARIANT
<222> (3)...(3)
<223> The K at amino acid position 3 can also be R.

<221> VARIANT
<222> (1)...(4)
<223> Xaa = Any Amino Acid

<400> 11
Arg Xaa Lys Arg
1

<210> 12
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Zinc binding motif of ADAM-TS-1.

<221> VARIANT
<222> (1)...(5)
<223> Xaa = Any Amino Acid

<400> 12
His Glu Xaa Xaa His
1 5

<210> 13
<211> 11
<212> PRT
<213> Artificial Sequence

<220>
<223> Conserved motif found in matrix metalloproteinases
and ADAMs.

<221> VARIANT

<222> (1)...(11)

<223> Xaa = Any Amino Acid

<400> 13

His	Glu	Xaa	Xaa	His	Xaa	Xaa	Gly	Xaa	Xaa	His
1				5					10	

A. CLASSIFICATION OF SUBJECT MATTER IPC 7 C12N15/57 C12N15/63 C12N9/64 C07K16/40 C12Q1/37		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols) IPC 7 C12N		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the International search (name of data base and, where practical, search terms used) EPO-Internal, BIOSIS, STRAND, EMBL, WPI Data, MEDLINE, EMBASE, CHEM ABS Data		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HURSKAINEN TIINA L ET AL: "ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of zinc metalloproteases: General features and genomic distribution of the Adam-TS family." JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 274, no. 36, 3 September 1999 (1999-09-03), pages 25555-25563, XP002158991 ISSN: 0021-9258 see ADAM-TS7 abstract; figure 1C figure 2C --- -/--	1,6-11, 13-32
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C. <input checked="" type="checkbox"/> Patent family members are listed in annex.		
* Special categories of cited documents : "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. "&" document member of the same patent family		
Date of the actual completion of the international search 31 January 2001		Date of mailing of the international search report 06.02.01
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016		Authorized officer Morawetz, R

INTERNATIONAL SEARCH REPORT

Intern	Application No
PCT/US	00/29380

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DATABASE PIR2 [Online] Accession number T14764, 20 September 1999 (1999-09-20) WAMBUTT, R. ET AL.: "Hypothetical protein DKFZp434H204.1 - human (fragment)" XP002158992 the whole document</p>	1
X	<p>--- DATABASE EMBL [Online] Accession number AL110226, 30 August 1999 (1999-08-30) WAMBUTT, R. ET AL: "cDNA DKFZp434H204 (from clone DKFZp434H204)" XP002158993 the whole document</p>	7-9
X	<p>--- BOR LUEN TANG ET AL: "ADAMTS: A novel family of proteases with an ADAM protease domain and thrombospondin 1 repeats" FEBS LETTERS, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, vol. 445, 26 February 1999 (1999-02-26), pages 223-225, XP002141413 ISSN: 0014-5793 the whole document</p>	1
X	<p>--- KUNO ET AL: "Molecular cloning of a gene encoding a new type of metalloproteinase-disintegrin family protein with thrombospondin motifs as an inflammation associated gene" JOURNAL OF BIOLOGICAL CHEMISTRY, US, AMERICAN SOCIETY OF BIOLOGICAL CHEMISTS, BALTIMORE, MD, vol. 1, no. 272, 3 January 1997 (1997-01-03), pages 556-562, XP002093741 ISSN: 0021-9258 cited in the application the whole document</p>	20
A	<p>--- -/--</p>	1-19, 21-32

INTERNATIONAL SEARCH REPORT

Interns

Application No

PCT/US 00/29380

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	VAZQUEZ FRANCISCA ET AL: "METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new family of proteins with angio-inhibitory activity" JOURNAL OF BIOLOGICAL CHEMISTRY, THE AMERICAN SOCIETY OF BIOLOGICAL CHEMISTS, INC., US, vol. 274, no. 33, 13 August 1999 (1999-08-13), pages 23349-23357, XP002142875 ISSN: 0021-9258 cited in the application	20
A	the whole document	1-19, 21-32
P,X	--- WO 00 58473 A (CURAGEN CORP ; LEACH MARTIN (US); SHIMKETS RICHARD A (US)) 5 October 2000 (2000-10-05) see SEQ ID NO:2285, SEQ ID NO:2286 page 1 -page 92 page 101 page 166 page 1672 -page 1680	1-23,25, 27,29-32
P,X	--- DATABASE SWALL [Online] Accession number Q9NPM2, 1 October 2000 (2000-10-01) AUFFRAY, C. ET AL: "ADAMTS7, alternatively spliced product (fragment)" XP002158994 the whole document -----	1

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 00/29380

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

Although claims 22-28 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.
2. ☒ Claims Nos.: 33, 34
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 33, 34

Present claims 33 and 34 relate to a product defined by reference to a desirable characteristic or property, namely an agent that interacts with a polypeptide of claim 1 or a nucleic acid sequence of claim 7.

The claims cover all products having this characteristic or property, whereas the application provides support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT for none of such products. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

Information on patent family members

intern:

I Application No

PCT/US 00/29380

Patent document
cited in search report

Publication date

Patent family member(s)

Publication date

WO 0058473

A

05-10-2000

NONE

